

Call for COVID-19 rapid response data science taskforce

Contents

Introduction	1
Research challenges	2
Funding	2
Who should apply?	2
1. Call for Existing Teams	3
Required skills, expertise and background	3
2. Talent pool	5
Ways of working	5
How to apply	7
Assessment	7
Outcome notification	8
Timeline	8
Contact points	8
Annex	9

Introduction

In 2019 a novel coronavirus-induced disease (COVID-19) emerged in Wuhan, China. A month later the Chinese Centre for Disease Control and Prevention identified a new beta-coronavirus (SARS coronavirus 2, or SARS-CoV-2) as the aetiological agent. Most of what we currently know of COVID-19 comes from data about its outbreak in China and Italy, but there may be differences in presentation and course within the multi-ethnic communities of the UK which have not been described thus far. Within the UK, high level mapping is being captured by Public Health England, but there is a lack of in-depth and granular data capture. This in-depth data is needed to inform prognostic prediction tools, enable effective clinical care or model the ongoing and likely future impact of COVID-19.

DECOVID is an emerging research collaboration between The Alan Turing Institute, NHS Trusts, HDR UK and other academic partners. This collaboration aims to use near real-time health data as the Covid-19 pandemic unfolds to allow researchers and clinicians to identify factors and generate insights that can lead to more effective clinical treatment strategies. DECOVID will be affiliated to PIONEER - the Health Data Research Hub for Acute Care.

We now wish to draw upon the broad network of researchers within the Turing community, to seed clinical and operational teams to support Trusts as part of this collaboration.

This is a rapid response call for teams of researchers as well as individual researchers to join us in addressing these urgent questions. Teams will rapidly trial tested, robust and reproducible data science and machine learning methods on this new national data resource of COVID 19 data, to create a near real-time, human-in-the-loop analytics platform and information console to support acute care management and clinical decision making. DECOVID has been structured to be nationally scalable.

Research challenges

To inform the call to action, urgent actionable questions are being developed in close collaboration with clinical leads. These questions will be defined and refined by clinical leads in collaboration with the Turing's Health Programme Director and a number of supporting senior colleagues in cross functional roles and across academic institutions.

Some examples of potential questions this project will seek to answer are included as an Annex.

Initially at least, teams will not conduct novel methodological research – but rather use tested, robust and reproducible methods to inform acute care management and clinical decisions.

Funding

We are aiming to secure funding to support project teams, potentially at 80% FEC or above. Researchers will not be expected to commit until we can confirm the funding. If researchers are able to commit their time without further funding, this is appreciated and you should note this in your application, but not expected and it is not a requirement for putting yourself forward.

Who should apply?

We are looking to on-board multiple UK-based data science teams, with strong experience in working with health and clinical data applications, to each conduct focused work on an individual clinical question as defined and refined by the clinical leads and Scientific Steering Committee.

Applications are initially sought from academics.

There are two parallel calls:

1. Existing teams – initial deadline Friday 3 April 13:00 BST
2. Talent pool – initial deadline Friday 10 April 13:00 BST

There are also two 'waves' to this project:

- Wave 1: Initially at least, we need teams not to conduct novel methodological research – but rather to use tested, robust and reproducible methods to inform acute care management and clinical decisions

- Wave 2: Conduct novel methodological research, aiming to improve on standard models (only once initial teams set up, system proven robust, and processes in place for validation – expected in a few weeks)

The application form will ask applicants to indicate whether the application is from a Team or for the Talent Pool, and whether you are interested in Wave 1 and/or Wave 2. If you indicate Wave 2, your application will not be immediately reviewed but will be put forward automatically when Wave 2 becomes active.

1. Call for Existing Teams

Due to the nature of remote work and the need to on-board teams swiftly and make progress, we have made the decision to **initially seek existing teams** who already work collaboratively and have a strong expertise match to the criteria to put themselves forward.

Existing teams that we seek may currently be university research groups or cross-university collaborations. We expect that their internal team structures may be broadly composed of the following roles:

- **Data Science Team Leader**, with experience leading distributed teams
- **Deputy Data Science Leader**, (particularly important, in case Team Leader is ill)
- **Data scientist team members** comprising Post-Doctoral and Doctoral students
- **Research software engineers** who develop robust, reproducible, reusable code and analysis pipelines

We welcome teams of other compositions; not all teams look the same. We will also look to augment and support teams with additional expertise from the Turing and other partners (as per the “Talent Pool” category of this call). Applications from diverse teams are strongly encouraged – diverse teams create breadth of thought and novel approaches to problem solving.

The Turing Research Engineering Group will also work with the teams – please see [Ways of Working](#) for further information on the role and function of the Turing’s central REG team on this project.

Teams in Wave 1 should be able to commit to this project ideally full time for at least the next three months and be able to start on or around 10 April.

The initial deadline is Friday 3 April 13:00 BST. The application form will remain open to continue to allow way of collecting details of any further interested teams who could be called upon later, or who may be interested in taking part in the ‘wave 2’ research.

Required skills, expertise and background

For all positions:

- Agile and flexible to changing/uncertain requirements
- Strong collaborative mindset and approach with ability to build strong working relationships across disciplines
- Commitment to help shape and deliver results and impact for the greater good, beyond traditional academic metrics
- A pragmatic approach to solving complex clinical problems with appropriate, established techniques

- Able to conduct project with minimal input from clinical colleagues
- Able to work remotely

For data scientists:

Essential:

- Proven knowledge in modern statistical methods and applications and machine learning, specifically *in one or more* of the following: statistical risk prediction, longitudinal analysis and modelling, uncertainty quantification, time series, survival analysis, multivariate methods, medical statistics, missing data, causal inference, feature selection etc, feature extraction, deep learning, classification, causal models etc
- Experience working with health data
- Experience working with real world, high-dimensional data
- Strong knowledge and experience of R or Python, cloud infrastructure and GitHub
- Experience and training in clinical data protection, management and governance – or willingness to rapidly learn and undertake training online
- MSc or PhD degree in computer science, statistics, mathematics, healthcare informatics, machine learning or equivalent

Preferred/desirable:

- Experience working with acute care data
- Experience in any of the following applied healthcare research areas: Epidemiology, operational research, forecasting, causal inference, prediction modelling
- Experience of analysing administrative datasets, particularly from the NHS including but not limited to electronic health records
- Experience in electronic health record data wrangling
- Experience implementing and evaluating ML algorithms in healthcare
- Experience in creating infographics / visualisations, and communicating uncertainty
- Experience of working within a Turing Safe Haven or other safe haven environments
- Previously taken part in a Turing Data Study Group or equivalent sprint-based analyses with new and diverse collaborators
- Previously contributed to Open Source projects and are familiar with the approaches encouraged by the Turing Way around reproducibility, openness and collaboration

For team leaders:

As above, and additionally:

- Experience leading a team
- Can provide statistical leadership and strong applied knowledge of clinical research principles and requirements of healthcare stakeholders
- Ability to identify and formalise key research problems in clinical decision making
- Can conduct rapid experimentation and deep scientific engagement for real-world delivery and impact
- Experience collaborating in cross-functional and geographically dispersed teams.
- Experience of writing data science reports and briefings that can be understood by a policy, managerial or clinical audience

2. Talent pool

As these teams make progress, we expect a requirement to augment those teams with additional skills and talent; therefore, we are also seeking individuals or small groups to put themselves forward into a “talent pool”.

The pool of talent may be a more flexible time commitment, depending on the skills, expertise, and resources required by the teams.

We would like to specifically invite individuals who have clinical experience who can flexibly support the translation between clinical and statistical questions.

We will work with the applicants to ensure that each team has a mix of skills, expertise and diversity needed to tackle the data scientific questions at hand.

Required skills and experience

The required skills and experience are the same as listed for the Teams above, *however experience directly in health applications is not an essential requirement for the Talent Pool applications.* We may seek additional skills and expertise as progress is made; this document may be updated and reissued accordingly.

The initial deadline is Friday 10 April 13:00 BST. The form will remain open after that date to continue allowing interested colleagues to put themselves forward, with a rolling review.

Ways of working

Programme structure

The programme is overseen by a governance and management structure comprising senior representatives from the scientific and clinical communities, in addition to representatives from NHS Trusts and patients and the public.

Clinical questions

Research questions will be continuously identified and prioritised by scientific and clinical leadership teams as necessary based on new data flows/availability and progress being made; or, moving to other questions if there isn't a possibility of solving the question in a useful timeframe. Teams should be able to respond and pivot where necessary.

All teams must be committed to sharing their progress openly and transparently, and in line with standards for the safe and responsible use of healthcare data.

Analysis environments

Work will be undertaken in scalable, flexible analysis environment, equipped with a wide range of data science tools and software, will be made available inside a secure research environment..

Turing Research Engineering Group

The Turing Research Engineering Group (REG) will provide additional support through developing a data analytics platform (the Turing Analytics Hub) that will be available to

research teams to rapidly prototype data science models, and through setting up software tools for data exploration and model validation to assess their predictive power.

The REG team will comprise a central team with experience of model validation, benchmarking, infrastructure building and testing. As a team, the REG will perform three primary functions:

- 1) Data management - Development of pipelines and scripts to ensure automated delivery of appropriate datasets to analytics teams on a regular schedule. This includes version controlling the datasets.
- 2) Model validation / Quality assurance - REG will build the infrastructure required to test models against unseen data. This infrastructure will also be used to replicate findings from the more private data the researchers are working with against the more sensitive lightly pseudonymised data.
- 3) Production - The models that are shown to work well in the validation pipeline will be put into production to make prospective predictions on future data yet to be collected. This work strand will automate monitoring and benchmark models' performances with new data.

In addition to contributing to the tasks above, each REG member will be assigned to and liaise closely with a data science/analytics team. In this role, their primary function will be to serve as 'bridges' to the central REG team. They will work with the analytics team to understand the problems encountered, and their feedback will be used to drive the data management, model validation and productions tasks as described above. In this manner, we will ensure that all issues are solved once across all teams to avoid effort duplication and that resources are shared effectively across all analytics teams.

If required, an initial output from the REG could also be a report/dashboard to describe the contents of the initial data dump. This report generation will be automated so that members of the team can receive updates about the new data as they are sent from the Trusts. The REG will take advice about the metrics that the teams would like to see on the report over the extent of project life cycle.

Characteristics of the overall DECOVID project:

- Focused and mission-led – to target and innovate for improved health outcomes, driven by clinical need
- A culture of openness, inter-disciplinary, team science, focused on delivery of results and optimised for “frictionless” engagement
- A “permeable” structure that allows for resources to be deployed and shaped around an iterative rapid cycle of learning
- Considerate and sensitive to frontline staff in order to ensure this work is burden-free
- Nationally collaborative to draw on expertise from a UK-wide talent pool of data scientists and artificial intelligence researchers to join and engage with the mission.
- Agile teams in both their research and research engineering practices, building upon a track record of success in rapid prototyping.
- Expert critical knowhow of utilising and linking diverse, large scale health and care data resources of varying data modalities. Combined with data wrangling expertise necessary to make complex health and care data analysis “ready for research”, within the rapid iterative prototyping approach.
- Build upon established partnerships with key stakeholders across the ecosystem with a model for expanding to new partnerships as the need and opportunity arises.

- Promoting best practice in reproducible research, open science, and to provide published results and working practices to the benefit of the wider research community.

How to apply

Please submit details via a short Expression of Interest (EOI) form online via the Turing's [Flexigrant](#) site by **Friday 3 April 13:00 BST**

- Teams submissions should have all the information above collated for each team member and submitted by the Team Leader.
- Applications to the Talent Pool should be submitted by the individuals, or if a small group is applying, a designated lead applicant should submit the application to the Talent Pool.
- Please ensure your Head of School, PI / Group Lead / equivalent would approve you switching to one of these projects, if you were called upon.

The following information will be required:

- Applicant/s details
- Expertise and key skills as per criteria above
- Brief summary of expertise, particularly if significant and strong experience match to the criteria (essential and desirable), highlighting anything particularly relevant to the potential research questions or in which you/your team have unique skills and experience
- If applicable, highlight any resources that could be rapidly enjoined to the project, such as additional data, deployable software, etc.
- Link (or links) to evidence the technical skills stated in the previous question: up to 5 outputs relevant to this programme that are representative of your work (e.g. papers, code, reports)
- Confirmation of ability to commit full time (Teams only)
 - Less than 3 months with / without funding
 - 3-4 months with / without funding
 - 5-6 months with / without funding
- Indication of time commitment which could be contributed (Talent Pool only)
- Indicate which research questions most relevant to expertise
- If funding is required, approximate costs (please indicate 100% FEC costs)
- Any additional comments

Assessment

Applications will come into the Turing centrally via Flexigrant.

Applications will be assessed against how strongly they meet the skills and expertise requirements described in the call above.

Existing teams

Stage 1:

EOIs from Turing university partners will be reviewed by the Turing University Lead or a Turing Fellow from that university designated by the TUL. (The TUL may put

together a small group to support them in this process.) The University will then put their top applications through to Stage 2.

EOIs from non-Turing university partners will be reviewed by the stage 2 panel.

Stage 2:

Short-listed applications will then be assessed by the Health programme team and the project team.

Talent pool

The review process for the Talent Pool will be set up in the coming week prior to the initial deadline of 10 April.

Outcome notification

Stage 1 outcomes will be communicated by the university to the applicant/s.

Stage 2 outcomes will be communicated by the Turing.

We will get back to applicants as soon as possible with the outcome.

Due to the high levels of interest we expect we will get many more offers from excellent teams than we will be able to appoint. We anticipate that we may be able to expand out teams later so may come back to applicants in the coming weeks.

Timeline

Opening date	Wednesday 1 April
Closing date	Friday 3 April 13:00
Review period	3-8 April
Outcome notification	9 April
Project start date	10 April or thereafter

Contact points

If you have any queries, please contact the team via covid19@turing.ac.uk

What we will do with your information

The personal information that you provide within the application will specifically be used for administering this call and the projects. The information will be viewed by those administering the call and involved in the collaboration (which may include Turing Fellows, staff, and researchers and staff from the partners), and your information will not be used for any other purpose without your specific consent.

Annex

Research challenges

– potential questions indicative of real questions which will be collated and prioritised by the Scientific Steering Committee:

- Diagnosis
 - What are the atypical presentations that have been processed as “non-COVID”, placing other patients and staff at risk?
 - What are the physiological biomarkers of poor and good outcomes?
- Triage and discharge
 - What features might be indicators of when to discharge patients?
 - Which types of patients require hospital care?
 - What types of patients are likely to re-present with respiratory failure?
- Deterioration
 - What features may indicate risk of deterioration?
 - How do features of risk differ between older and multi-morbid patients and that of young, healthy and healthy patients
 - When should patients receive ventilatory support?
 - Can we identify those at risk of catastrophic organ failure early?
- Operations and Monitoring
 - How can we better allocate resources across the Trust?
 - Can we understand variance in clinical practice?
 - Can we predict hospital demand based on NHS 111, online and primary care data?
- Impact on non-COVID patients
 - What is the impact of the COVID pandemic for non-COVID patients?
 - How is the restructuring of hospital services affecting elective services and outpatient visits?