Lancaster University

UCREL NLP Summer School 2024

Mini Hackathon and Team Collaboration

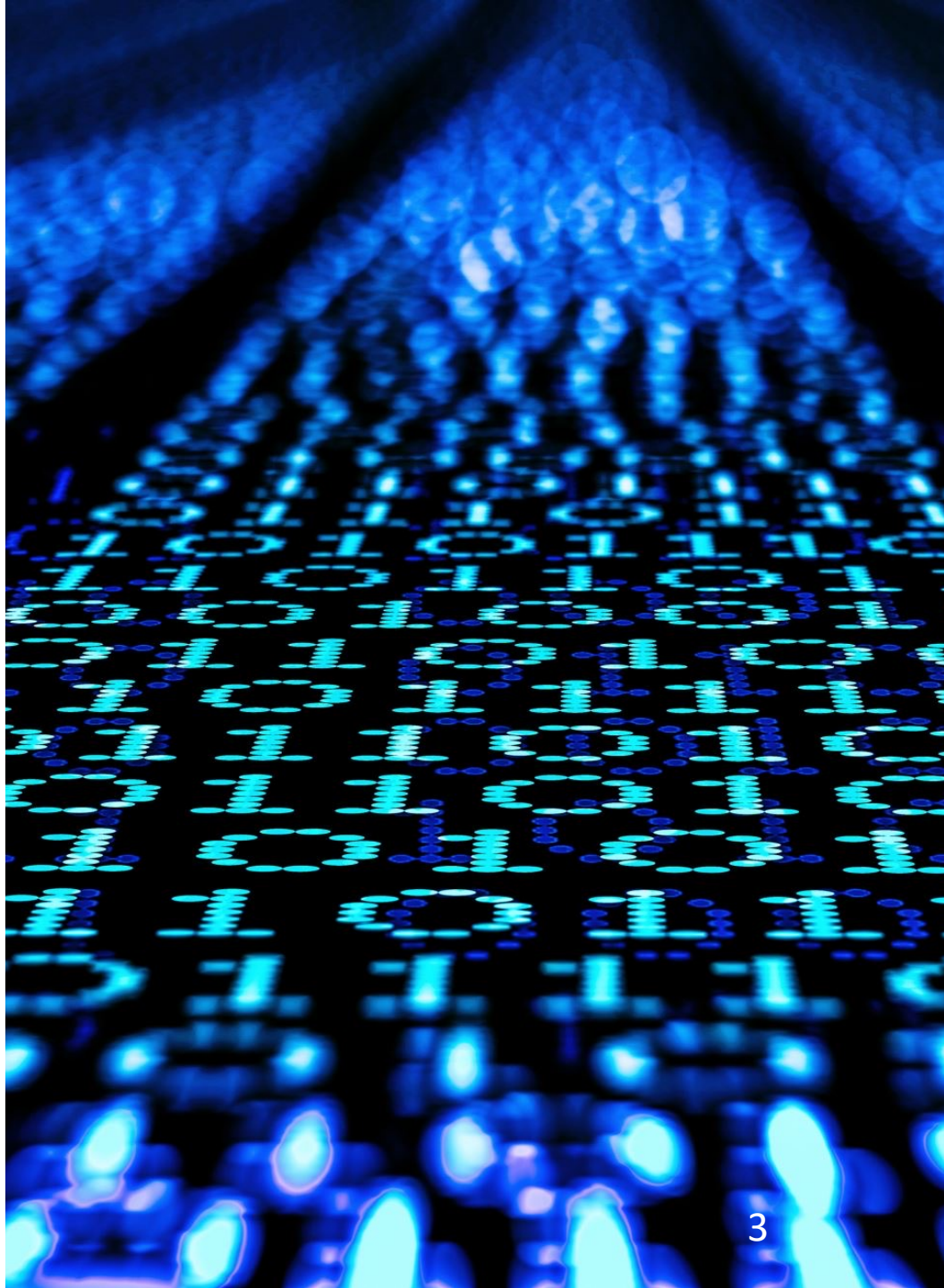
Paul Rayson, John Vidler, Daisy Lal, Nouran Khallaf, Frank Ratchakrit Arreerard, Scott Piao

Mini Hackathon and Team Collaboration

- Aims for these sessions
  - Further time to put into practice what you've started this week
  - Collaborate on a small(ish) task and produce some outputs
  - Consolidate your learning around a particular NLP challenge or approach
- Mentoring
  - Each team will have an assigned mentor from one of the UCREL team
  - There's no such thing as a silly question!

Inputs …

- All the corpus data and code that you've seen so far
- … your own previous research and data
- … any other online sources useful to you
- … all knowledge and skills you have learned, in this summer school and beyond

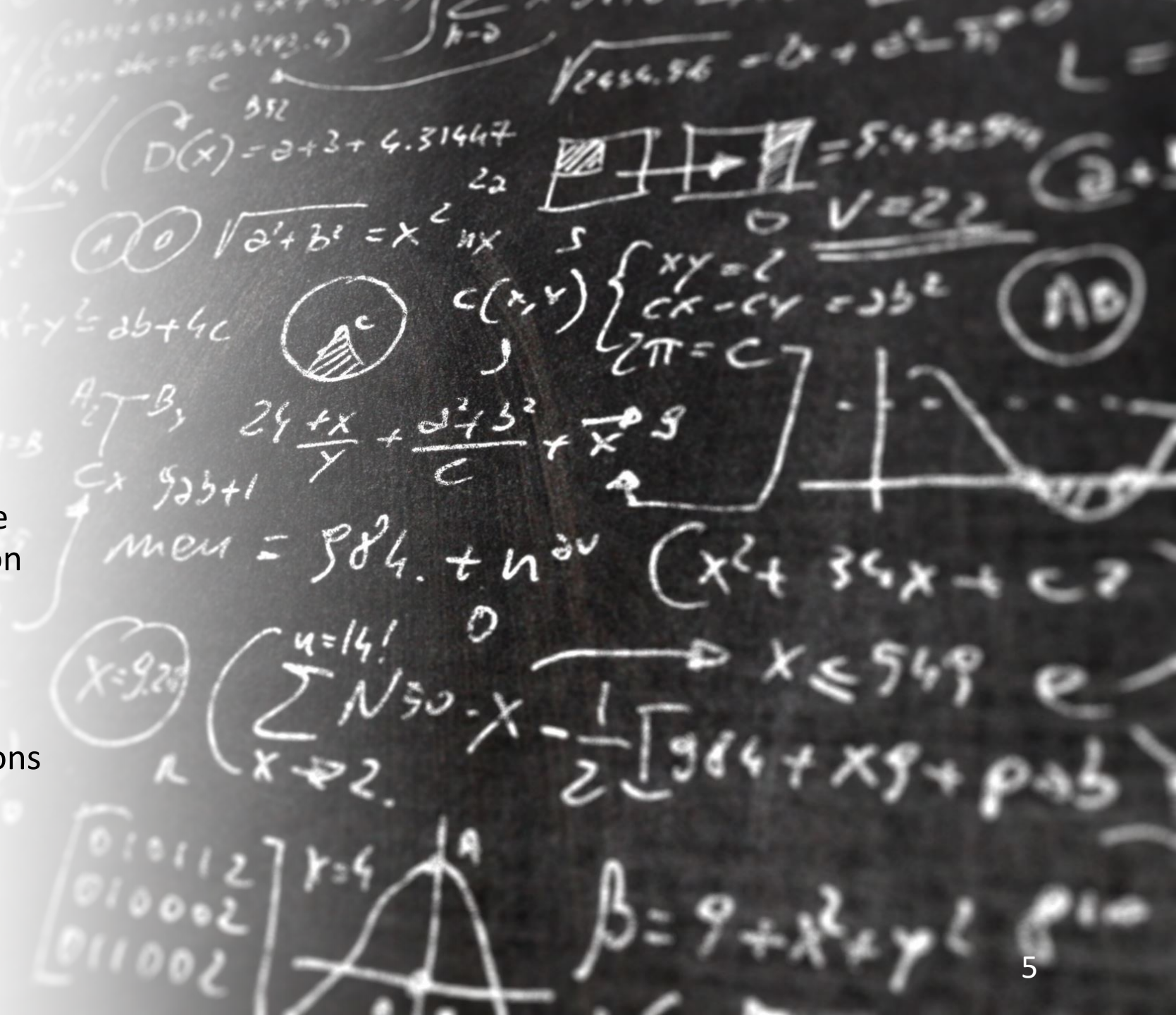


Suggested tasks …

- Take an existing NLP task and apply it to new data, evaluate it, improve it by training or fine-tuning
  - E.g. semantic tagging, emotion or sentiment analysis
- Take an existing dataset and define your own task, manually annotate it, modify the existing program code, train a classifier and replicate the manual tagging
  - E.g. (subsets of) GoEmotions, 4D reddits, ParlaMint, SemEval
- Research question?
  - How does this feature change over time?
  - Is this feature represented differently in subcorpora?
  - Is the data biased?
  - How to tune LLMs more effectively?
- Remember to follow …
  - Good practice guidelines for manual annotation
  - Sensible ethics approaches if you're collecting new data
  - Be aware the LLMs may produce biased outputs and pay attention to correct mis-information

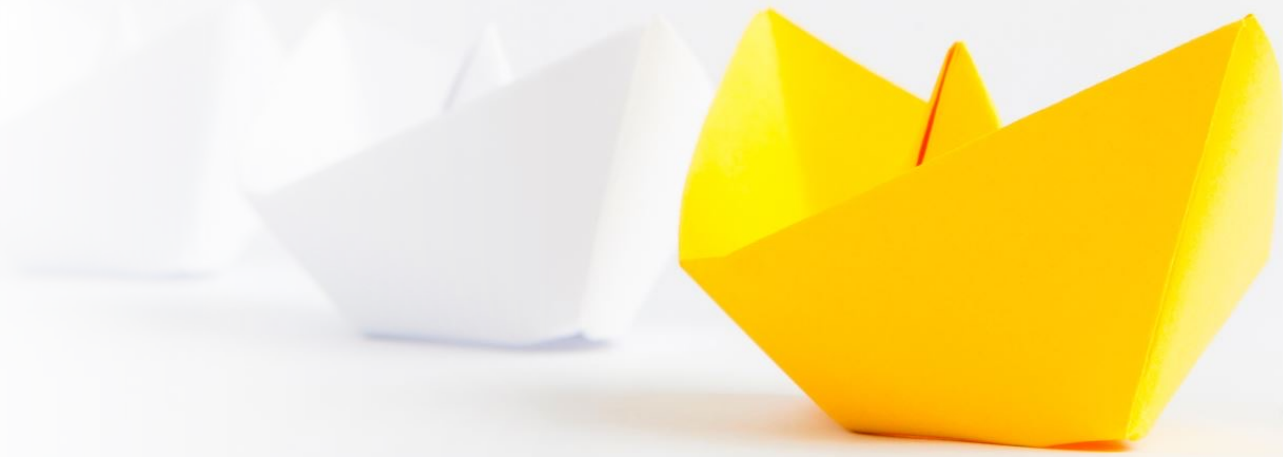
Outputs …

- A GitHub repository (can be private) with a readme describing what you've done
  - or equivalent – but please check with your mentor on the format
- A (manually or automatically) annotated corpus
- A set of comparative evaluations and experiments
- A Jupyter Notebook



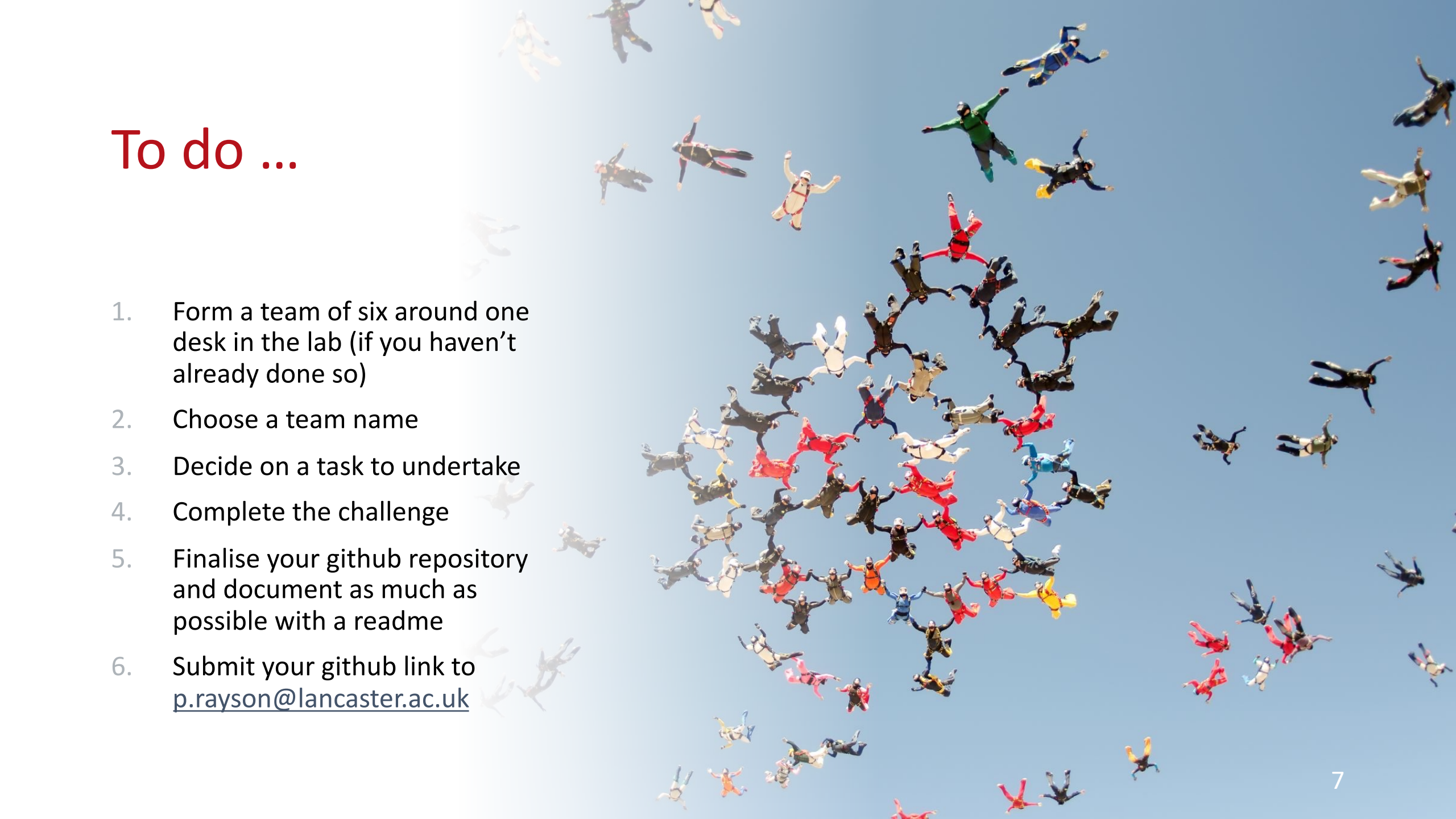
Evaluation criteria …

- We’re deliberately avoiding a leader board style comparison
- We’ll use a number of criteria to review the hackathon submissions
  - Language: Multilinguality, low-resource approaches
  - Novelty: Defining a new task?
  - Annotation: manual tagging methods or guidelines
  - Result presentation: visualisatons



To do …

1. Form a team of six around one desk in the lab (if you haven't already done so)
2. Choose a team name
3. Decide on a task to undertake
4. Complete the challenge
5. Finalise your github repository and document as much as possible with a readme
6. Submit your github link to p.rayson@lancaster.ac.uk



Go!

- Plan your time wisely, you only have 1 hour 45 minutes

Any questions?