Crafting Data-Centric Engineering NeurIPS 2021 Workshop proposal

---

In a nutshell

- *Organisers*: Dr. Benjamin Guedj (Inria and University College London, France and UK) [web] [email], Pr. Elizabeth Cross (University of Sheffield, UK) [web] [email], Dr. Christopher Nemeth (University of Lancaster, UK) [web] [email], Dr. Adam Sobey (University of Southampton, UK) [web] [email]

- *Format*: 1-day workshop, online

- *Features*: keynote talks, contributed talks (in particular from early-career researchers), panel discussion, poster session, networking

- *Website*: https://dce-neurips2021.github.io/dce-neurips/

- *Tagline*: Crafting the field of Data-Centric Engineering and Beyond

---

Brief outline. From autonomous vehicles and 3D printing through to smart cities and digital twins, the gap between our physical and digital worlds is growing ever smaller. At the interface of these two worlds is Data-Centric Engineering, a rapidly emerging new branch of science for the 21st century which combines the power and insight available from large-scale data sources with the tools and technology that shape our real-world environment. Bridging the gap between the digital and physical realms requires the development of new machine learning algorithms that are capable of processing large volumes of data from satellites, cell phones, distributed sensors, etc. and making robust and transparent decisions that improve our daily lives and protect our natural environment.

As Data-Centric Engineering is an emerging discipline it needs to define its boundaries, determine the techniques of importance and nurture a diverse community of ECRs to take these ideas forward. The focus of this workshop is to bring the engineering and machine learning communities together to define Data-Centric Engineering, through keynote talks and contributed sessions, panel discussions, poster presentations and networking.

Keynote speakers. Four 50 min talks from leading experts in data-centric engineering.

- Pr. Emily Fox, University of Washington and Apple, USA, [web] [tentative]

- Dr. Luc Julia, Samsung, USA [web] [**confirmed**]

- Pr. Pierre Pinson, DTU, Denmark [web] [**confirmed**]

- Pr. Brian Wardle, MIT, USA [web] [**confirmed**]

Contributed talks and poster session. Should the workshop proposal be accepted, we will run a call for papers from mid-July to mid-September, and aim for emailing notifications to authors by early October. We anticipate 50-80 submissions – papers must be four pages long maximum (excluding references). All accepted papers will be featured in the poster sessions, and we will invite authors of the best 6-8 papers to give contributed talks (20 min). All authors of accepted papers will be invited to submit full papers to the Data-Centric Engineering journal[1] after the workshop.

Panel discussion. We will run a 1h30 panel session to provide an opportunity for workshop participants to engage with experts. We very much hope to make this workshop a structuring milestone for the emerging field of data-centric engineering – as such, the panel will be mostly composed of early- and mid-career researchers, in an effort to give a voice to a rising generation of scientists. A summary of the panel discussion will be published after the workshop on the blog page of the Data-Centric Engineering journal.

- Dr. Rosella Arcucci, Imperial College London, UK [web] [**confirmed**]

- Pr. Mark Girolami[2] (Chair), University of Cambridge and The Alan Turing Institute, UK [web] [**confirmed**]

- Dr. Alexandre Lacoste, ElementAI - ServiceNow, Canada [web] [**confirmed**]

- Pr. Christophe Ley, Ghent University, Belgium [web] [**confirmed**]

- Dr. Alexandra [Sasha] Luccioni, Université de Montréal and Mila, Canada [web] [**confirmed**]

- Dr. Kimberly Tam, University of Plymouth, UK [web] [**confirmed**]

Diversity. The call for papers will be widely circulated, and in particular sent to the following groups: WiML, Black in AI, LatinXAI, Queer in AI. We will be particularly welcoming contributions from underrepresented groups in the machine learning and engineering communities. We will ensure gender balance among speakers and a diversity of backgrounds and affiliations.

Logistics. Estimated number of attendees: 300. This estimate is based on the number of papers accepted to the NeurIPS 2020 ML4Eng (https://ml4eng.github.io/) workshop (which is the closest to our proposal, among NeurIPS workshops organised in the past few years), where 59 papers were accepted and each paper had approximately 3 authors (accounting for about 150 attendees). We also anticipate a large fraction of the regular attendees (200-250) to the monthly Data-Centric Engineering webinar series[3] will also join (accounting for another 150 headcount). We anticipate a larger audience as the workshop announcement and call for papers would be largely circulated among the data-centric engineering community, including authors and supporters of the Data-Centric Engineering journal.

---

[1] https://www.cambridge.org/core/journals/data-centric-engineering

[2] Editor-in-Chief of the Data Centric Engineering journal, Director of the Data-Centric Engineering Programme at the Alan Turing Institute.

[3] https://www.cambridge.org/core/journals/data-centric-engineering/data-centric-engineering-webinar-series

Organisers. The organisers are all members of the Editorial Board (Associate Editors) of the Data-Centric Engineering journal (Cambridge University Press, https://www.cambridge.org/core/journals/data-centric-engineering, founded in 2020). They are all involved in the organisation of the monthly webinar series which averages 200-250 participants on various topics related to artificial intelligence, machine learning, data science and statistics, and the engineering sciences.

- Dr. Benjamin Guedj (University College London and Inria, UK and France) [web]

  Ben is a senior researcher with interests spanning theoretical machine learning, statistical learning theory, computational statistics and deep learning. His main expertise is on PAC-Bayes theory: Ben was a co-organiser (with Francis Bach and Pascal Germain) of the NeurIPS 2017 workshop "(Almost) 50 shades of Bayesian learning: PAC-Bayesian trends and insights", and he gave (together with John Shawe-Taylor) a plenary tutorial at ICML 2019 "A primer on PAC-Bayesian learning". He is the founder and scientific leader of the Inria London programme, a joint venture between UCL and Inria as part of the strategic scientific partnership between France and the UK. Ben has organised many scientific seminars and workshops, among which the annual conference of the French Statistical Society (about 600 attendees) in 2015 and of the French Mathematical Society (about 300 attendees) in 2018.

- Pr. Elizabeth Cross (University of Sheffield, UK) [web]

  Lizzy is a Professor of Mechanical Engineering at the University of Sheffield, her main research interests span the fields of structural health monitoring (SHM), machine learning and nonlinear system identification. Most of her research projects focus on the analysis of large datasets from monitored structures (aircraft, bridges, wind turbines), where she employs data-driven algorithms to extract useful information for health assessment. She currently holds an EPSRC Innovation Fellowship developing physics-informed machine learning for structural assessment.

- Dr. Chris Nemeth (Lancaster University, UK) [web]

  Chris's research is in the areas of computational statistics and machine learning, specifically Markov chain Monte Carlo methods and Bayesian computation. He has expertise in the areas of probabilistic modelling, times series and approximate inference. He is currently chair of the computational statistics and machine section of the Royal Statistical Society (RSS) and in this role he organises at least four workshops each year, including sessions at the annual RSS conference.

- Dr. Adam Sobey (University of Southampton and The Alan Turing Institute, UK) [web]

  Adam's research is in the area of AI with applications to the marine sector. This includes evolutionary computation, learning in continuous environments and machine learning that fits to the ground truth, even when we don't know what that is. He has organised 3 national workshops on AI and data related topics in the last year.

Programme Committee. We have drawn a programme committee largely from the Editorial Board of the Data-Centric Engineering journal – which is motivated by the perfect alignement of expertise, and the fact that authors of accepted papers will be invited to submit full papers to the journal after the workshop, ensuring positive follow-up for participants and nurturing and expanding the network of active members of data-centric engineering. All members mentioned here have agreed to participate.

1. Pr. Mark Girolami, University of Cambridge and The Alan Turing Institute, UK [web]

2. Pr. Stéphane Bordas, University of Luxemburg, Luxemburg [web]

3. Pr. Roger Ghanem, University of Southern California, USA [web]

4. Dr. Detlef Hohl, Royal Dutch Shell, UK [web]

5. Pr. Youssef Marzouk, MIT, USA [web]

6. Pr. Omar Matar, Imperial College London, UK [web]

7. Dr. Sumeetpal Singh, University of Cambridge, UK [web]

8. Pr. Kenichi Soga, UC Berkeley, USA [web]

9. Pr. David White, University of Southampton, UK [web]

10. Pr. Philip Whithers, University of Manchester, UK [web]

11. Dr. Antonio del Rio Chanona, Imperial College London, UK

12. Luca Magri, University of Cambridge, UK

13. Pradipta Maji, Indian Statistical Institute, India

14. Steve Waiching Sun, Columbia University, USA

In addition to the above members, the following members will be called upon depending on the final number of submissions. We will ensure that all papers receive three reviews, and that no reviewer shall have to evaluate more than three submissions.

1. James L. Beck, California Institute of Technology, USA

2. Andrew Blake, Samsung Research Cambridge, UK

3. Ruth Boumphrey, Lloyds Register Foundation, UK

4. Eleni Chatzi, ETH Zürich, Switzerland

5. Wei Chen, Northwestern University, USA

6. David Gerber, Arup, UK

7. Omar Ghattas, University of Texas, Austin, USA

8. George Em Karniadakis, Brown University, USA

9. Lord Robert Mair, University of Cambridge, UK

10. Rob Miller, University of Cambridge, UK

11. Charles Wang Wai Ng, Hong Kong University of Science and Technology, Hong Kong

12. Michael Ortiz, California Institute of Technology, USA

13. Sankar Pal, Indian Statistical Institute, India

14. Karen Willcox, University of Texas Austin, USA

15. Zhishen Wu, Ibaraki University, Japan