REQUEST FOR QUOTE for an ethics review of machine learning in children's social care

What Works Centre for Children's Social Care

1. Background to What Works Centre for Children's Care

The What Works Centre for Children's Social Care has been created to foster evidence informed practice in the children's social care sector.

The Centre seeks to address two key challenges:
● Creating a better evidence base – by sponsoring new practice-focussed research in areas where existing research is absent, of inadequate quality, or lacking relevance to decision-makers.

● Ensuring that the Centre's work results in change, not just knowledge – by investigating and addressing the barriers that might otherwise prevent practice leaders and practitioners from using insights from evidence to inform their decisions.

2. Background to the Project

Machine learning is being increasingly used in fields traditionally guided by expert decision-making, including work in safeguarding children. This has ranged from building decision aids to assist experts to full-scale predictive models. Hence, the What Works Centre for Children's Social Care (WWC CSC) would like to commission a tailored review into the question: *"Is it ethical to use machine learning approaches in children's social care systems and if so, how and under what circumstances?"*.

Previous work has largely occurred in a US context, see for instance this general review of US work by the Department of Health and Human Services[1]. However, there has also been recent pilots of work in using predictive models in safeguarding in the United Kingdom, for example in the London Borough of Hackney by Xantura[2] and in another county council by the Behavioural Insights Team[3]. The WWC CSC will be separately conducting pilot work with local authorities to improve our understand of machine learning approaches and their limitations[4].

3. What is being commissioned through this Request for Quote

In this review, we would like the successful researcher(s) to consider the ethics of the use of machine learning in a CSC context, via reviewing the literature and consultation of relevant experts and ethical frameworks. We are focussing in this review on **ethics** rather than acceptability in the sector or legal issues (except insofar as they overlap), which we will address separately.

We expect both a review of existing ethical frameworks, as well as recommendations specifically relating to the work the Centre will be undertaking (detailed here[45]), which relates primarily to the ethics of *researching* ML rather than its *deployment* in decision tools (except insofar as the latter informs the former). We anticipate that the key outputs of this work

---

[1] https://aspe.hhs.gov/predictive-analytics-child-welfare
[2] https://www.xantura.com/focus-areas/childrens-safeguarding
[3] http://www.behaviouralinsights.co.uk/publications/using-data-science-in-policy/
[4] https://whatworks-csc.org.uk/whats-new/blog/why-is-the-what-works-centre-researching-machine-learning

would be a 15-20 page report with recommendations with an accessible executive summary suitable for a blog post.

Particular questions to consider are as follows (though we appreciate that some of these are related), with more information on these trade-offs available in a report by the Nuffield Foundation here[5]:

1. For which decisions would it be ethically acceptable to use machine learning to assist social worker decision making, if anywhere?

2. Is it ethically acceptable to use ML on individual cases or only to make general decisions on broad groups? How should we consider tradeoffs between privacy and transparency?

3. What ethical or technical safeguards would need to be put in place, both thinking about the use of data and how digital tools based on machine learning would be constructed?

4. What kind of data would be ethical to use: for example structured, administrative data or text data, likely in the form of case notes?

5. Is it ethical to use models that may be complex or use nonlinear patterns or variable interactions to make decisions or assist predictions? How should we consider tradeoffs between accuracy and fairness?

6. Which interpretability frameworks (e.g. LIME[6] or other frameworks discussed here[7]), if any, suffice to provide transparency on how predictive models might operate? How should we consider tradeoffs between accuracy and interpretability?

7. Which ethical frameworks (e.g. FAT ML[8] or DCMS' framework[9]) represent best practice in considering general ethical principles for applying machine learning in children's social care specifically?

8. How can we guard against the kinds of biases relating to protected characteristics that have been a problem in predictive policing applications, including as detailed here[10]? How should we consider tradeoffs between accuracy and fairness?

4. Milestones and Indicative Budget

Milestones
- Work commenced by 31st March
- Draft report submitted by 30th August
- Final report submitted by 30th September
- Final report published by 14th October

Budget
The proposed budget for this review is up to £20,000 including all expenses and travel costs (excluding VAT).

---

[5] http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf

[6] https://homes.cs.washington.edu/~marcotcr/blog/lime/

[7] https://christophm.github.io/interpretable-ml-book/

[8] http://www.fatml.org/

[9] https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

[10] https://arxiv.org/pdf/1805.05859.pdf

5. Indicative Timetable

Set out below is the proposed procurement timetable. This is intended as a guide only.

Activity	Dates
Proposal deadline	Midday, 11th March
Clarification calls, if necessary	14th March
Successful organisation notified	Week beginning 18th March
Contracting must be complete & work commenced	31st March

6. How to apply

Please answer all questions in Annex 1 of this RFQ. Your entire proposal including all supporting documentation should be e-mailed to wwccsc@nesta.org.uk by no later than Midday GMT, on Monday 11th March 2019.

Applicants should be able to demonstrate:

- Experience in conducting literature reviews or systematic reviews.
- Experience in conducting reviews into applied data ethics, including in issues surrounding privacy, the use of data and issues concerning bias against minority groups.
- Experience in consulting with sectors on the use of data, including with relevant academics.
- Experience in writing reports which are concise and accessible to non machine learning experts / non data ethics experts (and ideally, social workers or other professionals).
- Experience in research into the interpretability of machine learning models.
- Experience in the children's social care system in the UK desirable, but not essential. Partnership with experts in the CSC sector would provide a significant advantage, but we can facilitate access to social workers and/or academics (through open invitation), if required.

If you have a specific question not answered either above or in the Annexes to this RFQ, please email louise.reid@nesta.org.uk.

Annex 1

Selection questions

1. Please detail your organisations relevant experience of completing similar projects and provide up to 3 specific examples. In each case you should state a) the work undertaken, b) the timescale for completion, and c) how you dealt with issues that arose, d) the value of the work. [Max 250 words]

2. Please describe the relevant experience of the individuals you propose to complete the project. [Max 250 words]
You may include up to two CVs for relevant individuals as appendices.

3. Describe your approach to the review - providing details of both the methods you will employ, how you will search the literature and how you will consult relevant experts. Further, describe your approach to writing the report and ensuring the language/framing is accessible to different audiences [Max 500 words]

4. Please provide a full breakdown of your proposed fees and pricing structure including the number of days, or estimated percentage of time, each team member is likely to spend on the project.

5. Please confirm if you are aware of any potential or actual conflicts of interest which may prevent you from working with the WWC CSC on this project. IF you consider a conflict may exist; please provide details in your response.

Annex 2

Terms and Conditions

Please note that the WWC CSC is currently hosted by Nesta; Nesta terms and conditions therefore apply to this submission.

This appendix provides the terms and conditions which will govern your submission of a proposal to Nesta. You are required to read this appendix carefully together with the Request for Quote (RFQ) brief (including all its annexes) before submitting a proposal. By submitting a proposal, you accept these terms and conditions and you agree to abide by

them. If you do not agree to these terms and conditions please do not submit a proposal.

1.   Submitting a proposal

1.1   By submitting a proposal you confirm that:

1.1.1   you have no actual or potential conflict of interest with Nesta (including Nesta's Trustees, officers or employees). If at any time during the RFQ process you discover an actual or potential conflict of interest, please inform Nesta promptly;

1.1.2   you have legal capacity to submit a proposal in response to this RFQ and are acting lawfully, ethically and in good faith in your dealings with Nesta;

1.1.3   you have sufficient time, skill, experience and resources to carry out the services to the highest professional standards expected of a competent supplier of services identical or similar and are able to obtain all necessary rights, licences, consents, waivers, approvals, permissions, permits, certificates and insurances necessary to provide the services to Nesta;

1.1.4   your quote will remain open for acceptance for a period of 90 days. A quote valid for a shorter period may be rejected;

1.1.5   all information contained in your proposal is true, accurate and not misleading; and

1.1.6   Nesta may share your proposal with any third party as Nesta in its absolute discretion deems necessary for the purpose of evaluation.

2   Compliance

2.1   Nesta reserves the right to reject or disqualify your quote where:

2.1.1   you fail to comply with the requirements of this RFQ (including but not limited to these terms and conditions), your quote is incomplete, or you are guilty of a serious misrepresentation in supplying information in response to this RFQ;

2.1.2   your quote is received after the deadline set out in this RFQ. Nesta will not consider any requests for an extension of the time or date fixed for the submission of responses;

2.1.3   there is a change in your identity, control, financial standing or any other factor impacting on the selection and/or evaluation of your quote;

2.1.4   you are or become insolvent or have a petition issued against you;

2.1.5   you do not have the economic and financial standing and/ or the technical and professional ability to carry out the services;

2.1.6   you are suspected either directly or indirectly of behaving in a collusive, canvassing, or anti-competitive manner or you offer or accept an inducement or reward in order to gain a

commercial, contractual, regulatory or personal advantage; and/or

2.1.7   you (or if you are a commercial entity, a director or person who has the power of control or power to make representations or decisions on your behalf) have committed any offence relating to conspiracy, corruption, bribery, fraud, money laundering or any other criminal offence related to your course of business or profession.

2.2   Nesta reserves the right in its absolute discretion to:

2.2.1   refuse any quote submitted;

2.2.2   disqualify any potential applicant who has been convicted of any of the offences listed at Regulation 23(1) of the Public Contracts Regulations 2006;

2.2.3   extend the time or date for fixed submission. In such circumstances Nesta will endeavour to notify all applicants of any change;

2.2.4   amend any aspect of this RFQ (including but not limited to the timeline) or cease the process at any time;

2.2.5   negotiate the award of additional services which are a repetition of the services advertised in this Request for Quote to the successful bidder;

2.2.6   limit the number of proposals invited to participate in any follow-up activity (for example, interviews) or to dispense with any follow-up altogether. Meetings will take place in London at our offices; and/ or

2.2.7   require that you clarify your quote in writing and/ or provide additional information and/or adequate references to Nesta's satisfaction. A failure to respond adequately may result in you not being selected.

3   Applicant Responsibilities

3.1   You shall be responsible for all of your own costs, expenses and losses which may be incurred in relation to the preparation of your quote, attendance at interviews, site visits or similar.

3.2   You shall at all times treat the contents of Nesta documentation as confidential, as well as any information regarding Nesta imparted to you by any other means, and only disclose such information as may be necessary for the preparation of a compliant response. At Nesta's request you shall return or destroy all documents, other materials, working papers relating to this RFQ and all copies thereof including all electronic copies. When completed you shall confirm such to Nesta in writing.

3.3   You shall not before the date and time specified within the RFQ documentation disclose to any person the amount of your proposal except where the disclosure in confidence is necessary to obtain insurance premiums or guarantees required as part of any proposal to Nesta.

3.4   Any contract entered into as a result of this RFQ will be on Nesta's standard terms and conditions (copy available on request).

4   Intellectual Property

4.1   All intellectual property rights in this RFQ and all materials provided by Nesta or any third party acting on its behalf shall remain the property of Nesta.

4.2   Any intellectual property arising out of the provision of the services shall belong absolutely and exclusively to Nesta.

5   Warnings/Disclaimers

5.1   Nothing contained in this RFQ or any other communication made in respect of it between Nesta or its representatives and any party will constitute an agreement, contract or representation between Nesta and any other party. For the avoidance of doubt, receipt by you of this RFQ does not imply the existence of a contract or commitment by or with Nesta for any purpose.

5.2   The information contained in this RFQ does not purport to contain all the information which you may require. While Nesta has taken all reasonable steps to ensure, as at the date of this RFQ that the facts contained in it are true and accurate in all material respects, Nesta does not make any representation or warranty as to the accuracy or completeness or otherwise of this RFQ.

5.3   Nesta accepts no liability to you whatsoever and however arising and whether resulting from the use of this RFQ, or any omissions from or deficiencies in it.

6   Waiver

Failure or neglect by Nesta to enforce at any time any of the provisions of these terms and conditions shall not be construed nor shall it be deemed to be, a waiver of our respective rights hereunder, nor in any way affect the validity of the whole or any part of this agreement, nor prejudice our respective rights to take subsequent action.

7   Jurisdiction

This RFQ shall be governed by and interpreted in accordance with the laws of England and Wales and subject to the exclusive jurisdiction of the English courts.