

ETHICS REVIEW OF MACHINE LEARNING IN CHILDREN'S SOCIAL CARE: EXECUTIVE SUMMARY

January 2020





What Works for Children's Social Care

Acknowledgements

The authors of this report would like to thank What Works for Children's Social Care for commissioning this project and for providing continuous support as we carried out our research and writing. In particular we would like to acknowledge Michael Sanders, Louise Reid, and Vicky Clayton for their patience and commitment to making a difference. We are also incredibly grateful to the family members with lived experience of children's social care who shared their views and experiences with us and to the Family Rights Group for making this outreach possible and for helping us ground our research in the reality of social work practice. Lastly, we are thankful for the stakeholders who shared their experiences and for the support for ethical innovation that we found in local authority representatives, researchers, industry leaders, and social care and family rights interest groups.

Authors

Leslie D., The Alan Turing Institute, **Holmes L.**, Rees Centre, University of Oxford, **Hitrova C.**, The Alan Turing Institute and **Ott E.**, Rees Centre, University of Oxford.

Funding

Department for Education, England.

About What Works for Children's Social Care

What Works for Children's Social Care aims to seek better outcomes for children, young people and families by bringing the best available evidence to practitioners and other decision makers across the children's social care sector. We generate, collate and

make accessible the best evidence for practitioners, policy makers and practice leaders to improve children's social care and the outcomes it generates for children and families.

About The Alan Turing Institute

The Alan Turing Institute (the Institute) is the national institute for data science and artificial intelligence. Established in 2015, the Institute brings together thirteen UK universities. The Institute researches data science and AI applications to tackle some of the biggest challenges in science, society and the economy. The Institute's Public Policy Programme works alongside policy

makers to explore how data-driven public service provision and policy innovation might solve long running policy problems and to develop the ethical foundations for the use of data science and artificial intelligence in policy-making. The goal of the programme is to enable technology to have a positive impact on the lives of as many people as possible.

About the Rees Centre

The Rees Centre (the Centre) at the University of Oxford produces research evidence to improve policy and practice in the areas of children's social care and education. The Centre aims to improve the life chances and particularly the educational outcomes of those who are, or have been supported by children's social care services. In 2017, the Centre and the Thomas Coram Research Unit at UCL established the Children's Social Care Data

User Group which brings together academics, local authority data managers, analysts, charities and funders with a shared vision that administrative data from children's social care and other relevant agencies in England can be analysed and fed back into policy and practice to improve the way that children's social care services respond to children, young people and their families.

If you'd like this publication in an alternative format such as Braille, large print or audio, please contact us at: wwccsc@nesta.org.uk

FOREWORD

Whether you're a sceptic or an enthusiast, it is hard to avoid recognising that the use of machine learning in children's social care is growing, in some places very rapidly. For advocates of the tool, the use of advanced analytics has the potential to improve services and outcomes for young people and their families by helping to rapidly find patterns in complexity. For their opponents, these tools risk dehumanising families, ingraining patterns of discrimination, and compromising the professional judgement of social workers while increasing unwanted intrusion into family life.

Both groups are right. Used well, in some circumstances, there is little doubt of the power of these tools to help professionals to make positive changes. Used poorly, or in the wrong contexts, it has the potential to be useless, or to actively cause harm. The consideration of what "well" and "poorly" mean, and what the right and wrong circumstances are, is both a question of effectiveness - how well do the tools actually work - and one of ethics.

It is for this reason that I'm pleased that What Works for Children's Social Care has commissioned the Rees Centre at the University of Oxford, and The Alan Turing Institute, to conduct this review of the ethics of using machine learning in children's social care. By bringing a combination of rigorous academic research, a practical focus, and the thoughts and experiences of practitioners, professionals, researchers and families to bear, they have produced this report and the recommendations within it, which we are proud to publish.

This is not the final word on this topic, and it does not aim to be. Instead, I hope that it will contribute to a much needed debate and when and where machine learning is appropriate, and what safeguards need to be in place to ensure ethical practice.



Michael Sanders

**Executive Director
What Works for Children's Social Care**



EXECUTIVE SUMMARY

The promises and perils of machine learning in children's social care

There could not be a more important time to think about the role that ethics should play in the context of using machine learning (ML) technologies in the domain of children's social care (CSC). Across the press, academia, and the worlds of policy and practice, concerns abound about the possible impacts of the growing use of ML in CSC on individuals, families, and communities. Many express legitimate worries about how the depersonalising and de-socialising effects of trends toward the automation of CSC are harming the care environment and negatively altering the way frontline workers are able to engage with families and children. Others raise concerns about how these data-driven ML systems are merely reinforcing, if not amplifying, historical patterns of systemic bias and discrimination. Others, still, highlight how the mixed results of existing ML innovations are signalling widespread conditions of poor data quality and questionable data collection and recording practices.

While these trepidations are valid and are helping to sharpen society's focus on the salient ethical issues that most demand concerted attention, they perhaps tell only one side of a more complicated story. In less than a generation, the explosive growth of ML, and of applied data science more generally, has become a transformative social, political, and economic force the world over. By helping researchers, analysts and practitioners to identify and draw insights from complex patterns extracted from large datasets, ML models have found useful applications in bolstering evidence-based decision making across a growing variety of sectors from healthcare, education, and transportation to agriculture, energy, and environmental management. With its capacities to assist the public sector in improving the personalisation of services, the prediction and analysis of trends, organisational functioning, and resource allocation, ML technologies hold the potential to significantly advance public welfare and the social good.

Keeping both these promises and perils in mind, how then can society responsibly harness the immense salutary potential of ML innovation in the realm of CSC? Provided that the way to such an unlocking of ML potential could be found, using it to foster the safety, wellbeing, and flourishing of children in need and their families would be a compelling prospect. Such innovations could, for example, be used to craft interventions that safeguard the dignity of child and family alike by focusing on outcomes that

optimise family functioning, health, safety, and child development. They could empower families through the data-driven crafting of humane, informative, and strengths-based interventions that provide support for the achievement of their own self-defined goals. They could also provide insights at the organisational and institutional levels, improving the effectiveness and adeptness of service provision and providing empirical information for policy-formation.

The project at a glance

It is against this backdrop that What Works for Children's Social Care (WWCSC) requested this report on the research question "***Is it ethical to use machine learning approaches in children's social care systems and if so, how and under what circumstances?***"

The findings we present here take some preliminary steps to providing an answer. They are aimed at data scientists, policy makers, local authority (LA) children's services departments, civil servants, and citizens. Where possible, we have tried to avoid extensive technical discussions, and we have attempted, where necessary, to provide plain language definitions of specialised terms and background information to aid the non-technical reader.

This research is informed by a range of methods – a literature review, an integrative examination of existing ethical frameworks in social care and ML, a stakeholder roundtable with 31 participants, and a workshop with 10 family members who have lived experience of children's social care.

While our results are preliminary and still in need of further consultation, we offer, in what follows, a **three-tiered framework for thinking about the ethics of ML in CSC**. In order to make the ethical stakes and practical implications of the difficult and multi-level question posed to us by WWCSC as clear as possible, we have broken it down into three further ones around which these three tiers are organised.

The first tier asks: ***Should we be doing this?*** Here, we take an *external point of view*—a view from outside of existing practices of using ML models in CSC—which refrains from assuming that its use is legitimate *per se* so that the bigger picture issue of the very justification of that use can be tackled head on. The point here is to examine the ethical criteria that would make the use of ML in CSC justifiable if they were satisfied in real world settings and then to examine the problematic contexts in which such criteria might not, in actuality, be met. In this section, we bring together existing frameworks in the ethics of social work and the ethics

of machine learning and artificial intelligence (AI) in order to formulate an integrated ethics of ML in CSC. We then use these ethical criteria to consider whether there are empirical factors intrinsic to the wider system in which CSC is situated (including historical patterns of inequity, the context of austerity, and conditions of system, organisation, and participant readiness) which may prevent the justified application of ML in CSC.

The second tier poses the question: **Can we do this right?** It takes an *internal point of view*, which assumes that the use of ML in CSC can, in fact, be justified so that we can identify and explore responsible practices of ML innovation in CSC from the inside of the design and production of the technologies themselves and internally to their processes of implementation. In this section, we present standards for best practice across ML's design and deployment lifecycle, paying special attention at each step of the way to the CSC context.

The third tier poses the question: **What is to be done?** It takes a *forward-looking point of view* that is focused on the potential of data scientific insights to transform the future of CSC for the better. It fleshes out recommendations for optimising the capacity of future data scientific research, community- and family-based collaboration, and deliberate innovation intervention to produce tangible societal benefits and advance individual, familial, and public wellbeing.

First-tier findings

Primary among our findings in answering the first-tier question is the integrated ethical framework for the use of ML in CSC that we present in detail in the full report. In summary form, its basic elements are as follows:

Ethical values that set the direction of travel for the responsible use of ML in CSC

- Respect the dignity of individual persons, empower them, and value the uniqueness of their aspirations, cultures, contexts, and life plans
- Connect with each other sincerely, openly, and inclusively, and prioritise trust, solidarity, and interpersonal collaboration
- Care for the wellbeing of each and all, and serve others with empathy, selflessness, and compassion
- Protect the priorities of social justice and the public interest by ensuring equity, recognising

diversity, and challenging discrimination and oppression

Practical Principles that establish the moral justifiability of the integrated practices of social care and ML innovation

- Fair, sustainable, ever-improving social care
- Social care that supports and empowers
- Transparent, responsible, and accountable social care

Professional virtues that establish common principles of professional integrity shared by social work and responsible ML innovation

- Be sincere, honest, and trustworthy
- Uphold ethical values and best practices
- Lead by competence and example
- Maintain appropriate professional boundaries
- Make considered professional judgments
- Be professionally responsible
- Be objective and impartial in making professional judgments
- Use evidence-based reasoning when rendering decisions

Our ultimate aim in setting out an integrated ethics of ML in CSC is to put its resulting values and principles into an actionable form. Such a form should support practice and bring together all stakeholders involved in the complex, multi-level and collaborative processes of conceptualising ML applications and projects. It should help them to cooperatively define their objectives, and it should assist them in designing, deploying, and monitoring their applications responsibly.

What is needed for this is a vehicle of common commitment—a way for all those who are dedicated to doing good through the responsible design and use of data scientific applications to continuously coalesce around a mutual recognition of the ethical motivations, practical principles, and professional standards of conduct that should motivate, direct,



underwrite, and steer responsible practices of ML innovation in the field of children's social care. We will call this living document a **Commitment to care, collaboration, and understanding** and provide in the full report a preliminary mapping of what this might look like.

The final task we undertake in the first tier is to look closely at how the ethical values that lie behind the responsible use of machine learning in children's social care might provide a **critical yardstick** of sorts against which the application of this kind of technology in the sensitive and demanding domain of CSC can be measured. To do this, we consider several empirical factors, which might call into question the justifiability of using ML in CSC. In particular, we examine and analyse three such factors:

- Public management in the context of austerity
- System, organisation, and participant (SOP) readiness
- Social inequality and cycles of poverty and discrimination

Second-tier findings

In the second tier of this review, we respond to the question: **Can we do this right?** We investigate how the practical principles that we have articulated in the ethical framework might help to provide guardrails for responsible conduct. We also examine how such principles might give shape to best practices from a point of view internal to the boots-on-the-ground activities of ML innovation and use. To do this, we move step-by-step through the design and implementation pipeline of the production and use of ML models in CSC, paying special attention to domain-specific needs and potential pitfalls of the CSC use case. In outline form, here is what we cover in this second tier:

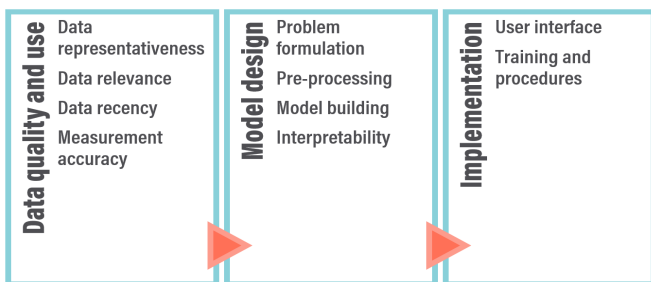


Figure 1. The stages of responsible machine learning innovation covered in this review

Third-tier findings

The final section of this review responds to the third-tier question, What is to be done? It provides some preliminary recommendations for steering the present direction of the use of ML in CSC, both in its application to practical, real-world problems and as a medium for research insight and discovery. It presents eight such recommendations:

1. **Mandate the responsible design and use of ML models in CSC at the national level.**
2. **Connect practitioners and data scientists across local authorities to improve ML innovation and to advance shared insights in applied data science through openness and communication.**
3. **Institutionalise inclusive and consent-based practices for designing, procuring, and implementing ML models.**
4. **Fund, initiate, and undertake active research programmes in system, organisation, and participant readiness.**
5. **Understand the use of data in CSC better so that recognition of its potential benefits and limitations can more effectively guide ML innovation practices.**
6. **Use data insights to describe, diagnose and analyse the root causes of the need for CSC, experiment to address them.**
7. **Focus on individual- and family-advancing outcomes, strengths-based approaches, and community-guided prospect modelling.**
8. **Improve data quality and understanding through professional development and training.**

While we conclude this review with recommendations, which are outlined in depth in the main report, we would also like to highlight from the outset that this study is primarily intended to help clarify some of the most substantial and complex ethical issues that arise in the context of the real-world application of ML in CSC. For this reason, the report should be utilised both as a means to reflect on external questions about the appropriateness and justifiability of using ML applications in CSC (both for specific use cases and in general) and as a preliminary guide for developing internal processes of data scientific innovation and implementation that incorporate ethics considerations at multiple points throughout the development and deployment lifecycle.



NOTES

NOTES

wwccsc@nesta.org.uk

[@whatworksCSC](https://twitter.com/whatworksCSC)
whatworks-csc.org.uk





What Works for
**Children's
Social Care**

wwccsc@nesta.org.uk

 [@whatworksCSC](https://twitter.com/whatworksCSC)

whatworks-csc.org.uk

