

AI Ethical Impact Assessment – Knowledge mining project

Responsibility exists across

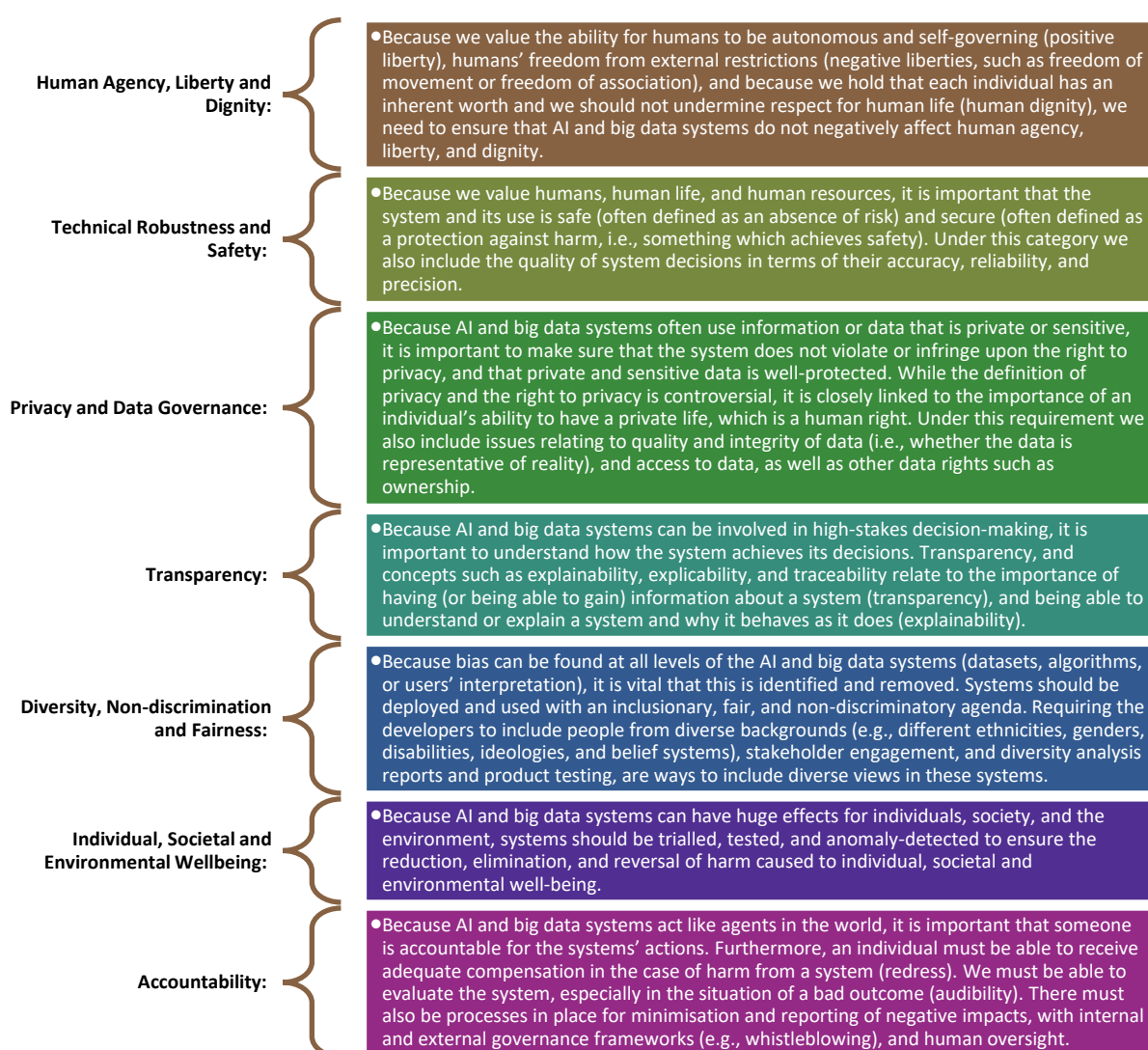
Technology – how the systems are developed and maintained, including managing and socialising the technical roadmap and associated opportunity.

Transformation – how the systems are integrated into new ways of working including managing impact assessments and cost/benefit analysis.

Services (Children and Young People's Services) – how the systems become cohesive parts of 'business as usual' and the monitoring and safety maintained.

With support from a number of other services such as Procurement, Legal, HR and Training.

Framework - For more information – see the NYC AI strategy (draft).



Name of Project	CHI10 – 3991 – Microsoft knowledge mining project (proof of concept)
Members of Ethics panel	<p>Project team – 6th November 2023</p> <p>Cath Ritchie Claire Wilson Mark Peterson John Jemson Jonny Hoyle Tom Moore Jessica Daly</p> <p>Reviewed by Project Board – 27/11/2023 Reviewed by Project Team – 14/3/2024 Updated by CR – 15/03/2024 Reviewed by CW and JH – 20/03/2024</p>

	Explanation and Discussion notes	RAG rating
1. Human Agency, Liberty and Dignity	<p>The tool can work to increase agency and improve freedom and dignity as it will map the networks around the child in a more child centred way. It will help find the neighbour who regular feeds the child or the football coach who has known them for years. It is not bound by family relationships only. This means it can help bring in a network of care of people that already care and love for that child and not break these bonds, as they often are in current ways of working.</p> <p>Legislation puts primacy on family bonds so there was a discussion as to if this tool will actually go against how we should work. It was agreed that this is an outdated way of thinking and there are plenty of examples now of courts being happy with the extended network providing care.</p> <p>Ultimately this tool helps support information retrieval and signposts to wider information sets – it provides breadcrumbs to the right information. The social worker will know and understand the child and make a holistic decision. The technology is informing the decision and does not rule in and out. This is congruent with social work practice and improves on the risks taken now, which is there is too much information to engage with quickly and effectively, so things will be missed.</p> <p>There can be weighting within networks – the tool highlights where there are more interactions. Can build on this using colours/coding etc. The way the information is displayed and visualised will impact decision making by the person and this must be designed around what works for social work and monitored to ensure it does not have over influence or bias.</p> <p>Risk is associated with data quality, error where solution finds or misses an entity (names, locations etc.) – this exists with any human looking at the information available (currently). Expectations are different in terms of human-error vs. machine-error. Consider the consequences of the tool missing information. Noting that data quality is poor. Mitigation is through strong practice model (which would incorporate specific guidance around this tool if it was used by practitioners).</p> <p>AI is not making a decision, it is looking at info and presenting it. It is used as a tool in a toolkit as a part of wider good practice. Enabling and supporting good social worker practice.</p>	Green at proof-of-concept stage.

	<p>Social worker's agency/dignity – does this tool undermine practice or is there a risk that skills and knowledge are lost? Agreed that this makes tasks easier ("head start") and doesn't replace important work/steps. Legal framework is unchanged and would need to be adhered to in the same way. Search functionality – should return results and the ambition is to be able to link back to full document/source data. Social workers will do check and balance.</p> <p>Training and qualified social workers use the Professional Capabilities Framework (PCF) and associated Knowledge and Skills framework. There is a responsibility on professionals to act with professionalism, ethical consideration and to adopt improved ways of working, which includes Digital Capabilities. This tool enhances working practice and elevates the use of data informed decision making. This speaks to the heart of social work as it arms the worker with detailed information to enable person centred decisions while leaning on their skills of critical thinking, analysis and risk management including ethical and safe practice. The tool also demonstrates the value of co-producing technology with social work leads, built around real-world practice so that it supports and enables rather than adding burdensome processes. Some may fear that using AI in this field is dangerous, but this discussion has surfaced that these professionals have exactly the right skills and standards to use AI effectively and appropriately.</p> <p>Professional standards - Social Work England Department for Education - (publishing.service.gov.uk) Professional Capabilities Framework (PCF) BASW</p>	
2. Technical Robustness and Safety	<p>Best practice data security has been implemented, all processing takes place behind private endpoints and within NYC tenancy, risk averse/industry standard approach. This work involved consulting with third-parties and NYC Data Gov and Information Security.</p> <p>Starting to work on testing the system (accuracy, reliability, precision).</p> <p>Some small risk as data moved to cloud (this is the corporate approach) – MS datacentre is UK South. MS security measures are robust. More info is in DPIA.</p> <p>No input mechanism from within tool, input has to be through source system – the tool be beneficial in terms of highlighting data discrepancies and could lead to an improvement in data quality.</p> <p>DPIA continually monitored and updated. DPIA approved by DPO.</p> <p>View that the biggest risk remains human and that the technology/robustness is the strongest it can be.</p>	Green at proof-of-concept stage.
3. Privacy and Data Governance	<p>DPIA</p> <p>There is nothing that the technology is doing that a human could not – the searching and networks/links could be made it would just take longer.</p> <p>Data is not being created, the data is already held by the council it is just being visualised in a different way. Visualisations are linked to source data.</p> <p>Potential challenges around data quality e.g., people known as different names or people requesting to be removed.</p>	Green at proof-of-concept stage.

	<p>It would be possible to look at data lineage to build rules/understanding. Rules applied and what is emphasised should be driven by what is valuable to a social worker, not what makes sense technically.</p> <p>LCS system currently flags risk factors e.g., a risky person or place and this could be carried over into the tool. A person may not want to be identified as a risk but practice dictates that they should be and the AI is not adding a risk or flag just pointing to it already being there. The social worker then engages and manages any actions.</p>	
4. Transparency	<p>DPIA – yes transparency, explainability, explicability and traceability, included as requirements but this needs testing and ongoing monitoring. We must ensure that users of this system understand what it is doing and are also transparent about their use of it to support their work.</p> <p>This should be done in a practical way with system users and be documented and shared as part of training and guidance. This tool allows for a greater level of transparency than the existing/as-is process. Co-produced with service users – but this may only involve a small cohort so would need to be communicated to all users. Training and guidance should reach all users whether occasional, agency or frequent users.</p> <p>Practice guidance being developed alongside tool.</p>	Green at proof-of-concept stage.
5. Diversity, Non-discrimination, and Fairness	<p>Equality Impact Screening Assessment completed and will remain live updated document as part of project approach.</p> <p>Principles of this point are aligned with social worker practice as a starting point, to protect and do no harm.</p> <p>Using AI could be more inclusive (to a degree) – searches for name, returns information without judgement. Currently no sentiment analysis is taking place on the data.</p> <p>However, from a tech point of view it will find names like John easier than other non-English names. AI solutions are biased in this way, any solution is only as good as the person who coded it and will adopt their bias. Some minority groups or languages may be deprioritised by MS for commercial reasons – expectation is that MS will be trying to be as inclusive as possible but we need to be vigilant to the impact of bias.</p> <p>Need to test and monitor in the real world to ensure we are happy and remain happy with the outcomes.</p> <p>Nicknames – could be picked up if context is understood, but this is more of an advanced solution, natural language understanding. Could be looked at and tested – e.g., probability scoring. Are certain words ignored by the model e.g., if they are considered offensive, but it could be a nickname? The team did work to validate ecomaps, through a manual test.</p> <p>Algorithmic bias – is a separate project risk (view is that MS have done work to mitigate this in developing their ‘off the shelf’ products) – it is a known risk with AI. We could identify it but could we expect to remove it (unlikely).</p> <p>Could build in/report on some items from the framework - requiring the developers to include people from diverse backgrounds (e.g., different ethnicities, genders, disabilities, ideologies, and belief systems). Stakeholder engagement, and diversity</p>	Green at proof-of-concept stage.

	analysis reports and product testing, are ways to include diverse views in these systems.	
6. Individual, Societal and Environmental Wellbeing	<p>System should be trialled, tested, and anomaly-detected to ensure the reduction, elimination, and reversal of harm caused to individual, societal and environmental well-being. Tests ongoing to ascertain a level of confidence based on the above.</p> <p>Rigorous testing and check and balance through project governance. Testing is from a starting point/viewpoint of challenging everything. Tests should be undertaken on an ongoing basis to monitor it, is it still ok? This should be built into any 'go-live' product lifecycle.</p> <p>Risk around tweaks/changes in models/mechanisms by MS not necessarily communicated. Things will be updated and we won't know. Included as a project risk.</p> <p>Ongoing testing needs to be done by experts with relevant experience to ensure high levels of confidence remain. At least annual checks.</p> <p>Environmental considerations have been picked up in Climate Change Impact Screening. Azure carbon neutral since 2012 more info Azure Sustainability—Sustainable Technologies Microsoft Azure</p>	Green at proof-of-concept stage.
7. Accountability	<p>In-line with policy/regulation (local, national, international).</p> <p>Draft EU AI regulation - the regulation would have extraterritorial reach, meaning that any AI system providing output within the European Union would be subject to it, regardless of where the provider or user is located.</p> <p>System would be auditable – who has used it, not clear yet about extent of audit - may not have granularity about searches, which records have been viewed.</p> <p>Importance of practice guidance to set into context, this is a tool in a toolkit, social worker decision-making still requires evidence and rationale behind decision making. This tool could not be used in isolation to make decisions. The social worker remains accountable for decision making. Guidance should not just be for end users but the wider ecosystem that supports them.</p> <p>Guidance should be accessible and co-produced with end users so that they can effectively manage the benefits alongside the risks.</p> <p>We are not passing any autonomy to the technology.</p> <p>Training and induction of social workers to include benefits, when should be used etc. feedback about the tool could be built into discussions at 121s or CPD. Could be useful to include case study, user story, examples.</p> <p>We have a duty of care to residents and employees – to be the most informed we can be and this will help that.</p> <p>Would need to monitor how the tool effected ways of working and any cultural shifts over time.</p> <p>There is not a risk that a person would accidently damage records as all the input of data is done on source system.</p>	Green at proof-of-concept stage.

	<p>Accountability for this data is well understood.</p> <p>Visualising data in this new way may actually help data quality as data quality issues will come to the fore more easily.</p>	
<p>NB</p> <p>Access to tool would likely be through a link or link to an app (through Teams), project is not looking at contextual linking (launch from LCS) – in theory can deep link back to documents/source data in LCS but not implemented or tested.</p>		
Recommendations from the panel	<p>CR and CW to write up and circulate. Done.</p> <p>Follow-up with Board review of this. Reviewed no amendments.</p>	
Governance route – project sign off	<p>Reviewed by Project Board – 27/11/2023</p> <p>Reviewed by Project Team – 14/3/2024</p> <p>Updated by CR – 15/03/2024</p> <p>Updated by CW and JH – 20/03/2024</p> <p>Approved by Project Board 22/04/2024</p>	