A. General recommendations:

1. Ground the impact assessment part on civil liberties rights in the existing UN Guiding Principles on Business and Human Rights, especially pillar II (corporate responsibility to respect human rights, including human rights due diligence and human rights impact assessments). Referencing the framework would already be helpful, in an effort to connect the framework to existing internationally recognized standards. In this way, the AI RMF can build on an existing framework, gain legitimacy and enhance harmonization and consistency across the industry, with human rights due diligence as a baseline.

2. Clarify that a separate methological tool – such as a model for impact assessment – will be needed, as this draft guidance is not sufficiently precise and ready for operationalization. At the European Center for Not-for-Profit Law (ECNL), we have mapped and analyzed various methods and models currently available (or under development) by a diverse set of stakeholders, from academics (e.g. Mantelero) to industry (e.g. Microsoft). The AI RMF could include and reference some of them to showcase this is already being discussed globally and across sectors. We would also recommend including minimum key criteria for a meaningful impact assessment model; ie. what are the minimum requirements necessary to make it effective. Please see a list of existing models already being proposed and/or piloted in Annex 1.

3. Propose levels for assessing the magnitude/severity and likelihood of risk, to provide an example of what risk levels could look like and what potential metrics could be involved (see e.g. Mantelero p. 31ss, for piloted model of risk scales and mapping).

4. Strengthen the requirement related to external stakeholder engagement.

B. Key risk and impact assessment features:

Key risk and impact assessment features	**Issues to Address**
Normative framework	Criteria, Procedures, Roles, Enforcement mechanisms
Scope	Type of impact(s), Benchmarks for harms and impact, Priorities
Timing	Stages and trigger points for assessment, Phases, Iterations
Methodology	Indicators for Impact, Scales for assessment, Metrics (quantitative and qualitative), Balancing and proportionality (trade-offs)
Engagement	Assessment teams composition, Identification of impacted stakeholders / groups, External experts, Frameworks for meaningful engagement
Oversight	Documentation and recording, Reviews, Responsible authority, Feedback loops, Indicators of performance
Accountability	Publication conditions, Monitoring, Assistance, Incentives

Content: criteria, indicators, benchmarks, methods

Risk assessments must include a holistic approach and assess the impacts of AI systems on a wide range of human rights, including collective rights, economic, social and cultural rights, and environmental rights, all serving in principle as benchmarks and criteria for the assessment. According to the UN Guiding Principles on Business and Human Rights, a credible human rights assessment should assess impacts on people, not impacts on business, draw on independent human rights expertise and highlight the concerns of affected stakeholders. In addition, based on the ongoing findings of increased bias and discrimination of AI systems, the assessment should prioritize harm reduction and the adverse human rights impacts on marginalized and vulnerable groups. Developing a standardised methodology for various AI contexts will be crucial to enable translation of human rights standards into the operational assessment process and clear quantitative and qualitative metrics. Methodology could be based on an analysis of the severity and likelihood of risks relevant to human rights – these should recognize the socio-technical context of the AI systems.

Process: when, how and with whom to implement, stakeholder engagement, public access

When conducting a impact assessment, procedural human rights must be respected alongside substantive rights; in particular: principles of non-discrimination, privacy, participation, empowerment and transparency. First, an impact assessment should be conducted at all stages of the AI lifecycle (starting with the ideation and design stage, through implementation and post-deployment evaluation), reviewing the impacts in an iterative manner.

Second, the process must center meaningful stakeholder engagement in the process. AI system developers need to engage with different stakeholder groups during the design and development of an AI system to deliver practical solutions of risk and harm mitigation. It's especially important to include diverse forms of expertise, lived experiences, and lessons from other sectors. Such processes should not be seen as a burden, but rather as an integral part of the usual AI lifecycle; it's possibly the only way to design rights-based and high-quality technology. This is not only a matter of regulation, but of a mix of social, institutional and organisational processes, intertwined in a truly interdisciplinary practice. Specific methods for such engagement are needed, particularly for vulnerable and affected communities. As a minimum, they should be included in the formulation of priorities, definitions, and outcomes of the impact assessment, and ultimately, in the decision whether and in which ways to deploy AI. There are different guidelines on civil participation and engagement in decision-making processes that can serve as a baseline for such models, and ECNL can offer one such engagement framework, currently in the final phase of development.

Third, appropriate resources and capacity must be allocated for this purpose to ensure adequate assessment. Impact assessment processes can be severely limited if conducted as a self-assessment exercise. To resolve this, impact assessments need to involve external assessors and/or oversight, and solutions needs to be investigated to make it economically sustainable, at least at a societal level (e.g. partial subsidization, engagement of civil society, research institutes, dedicated agencies, etc.). Finally, the findings of impact assessment processes should be made available to the public. This is an important step to enable public scrutiny, engagement, and potentially seek remedy in court for harmful deployments of AI. Clearly, the framework should include, where necessary, mechanisms to protect private trade secrets and intellectual property, but this can occur alongside providing access to key findings.

Outcome: oversight framework, enforcement mechanisms, accountability

Establishing an oversight framework for impact assessments to ensure they are conducted in a consistent and effective manner is critical, building upon different sources of law, layers and organisational procedures. The framework should include external, iterative, and ongoing review and oversight of the impact assessment, define what the relationship will be with compliance

programmes, and outline how the different enforcement approaches would relate to each other. All information related to an oversight body and its assessments or review should be made publicly available and accessible (note that this is a distinct transparency requirement than the one on the impact assessment process itself: for example, we may have a transparent impact assessment process, but an opaque oversight nonetheless). Furthermore, steps should be incorporated to enable ongoing feedback from the public on the effectiveness and impact of the framework, and revisions when necessary.

When AI is deployed by public institutions, these should develop and document indicators of performance of the technology, such as successes, failures, accuracy levels, purpose and outcome, to evaluate the ongoing or potential impacts of the technology in a holistic way. This also involves tracking the effectiveness of responses to impacts on individuals or populations that may be at increased risk of vulnerability or marginalisation. When AI is likely to affect certain communities in a disproportionate way, potential public consultation boards composed of individuals directly affected could be established and consulted. A smart mix of policy measures and different roles should be considered in an oversight framework for public authorities, standardisation bodies, private sector and interested public, to monitor, assess and revise AI impacts.

C. Framework for Meaningful Engagement of Stakeholders

The development of the Framework for Meaningful Engagement was a co-creation exercise led by ECNL for the past 8 months with over 130 participants from civil society, academia, law, policy, business sector and public institutions. Participants initially discussed the problems and barriers to engagement, and sought to capture broad ideas, existing knowledge and lessons from lived experiences from AI and elsewhere. This was then supplemented by desk research (in part crowdsourced from participants) by ECNL and SocietyInside. Engagement is not a new field and there are many examples of good practice in policy, industry, academia and civil society to draw from. The distillation of this knowledge to date is presented here for further deliberation with stakeholders during 2022, with the aim to pilot it in practice.

Numerous existing ethical guidelines, codes and "tech for good" commitments, as well as emerging national and international standards mention the importance of societal and stakeholder engagement in shaping the AI design development. But few give practical guidance on how to make this process meaningful. *That is the aim of this Framework.*

Here we hope to answer three essential questions for engagement:
1. What does 'meaningful engagement' actually mean?
2. What does a trustworthy engagement process look like in practice?
3. How can those involved distinguish the meaningful from the meaningless?

Please find the draft framework at this link and a presentation PDF attached as Annex 2.
https://app.mural.co/t/trustandcompanieswithgtl6355/m/trustandcompanieswithgtl6355/1657550136808/bc1851c060121ad7be72a799b60283a8e22a8904?sender=u06b407597dc11f94b09e3349

Annex 1: Impact assessment models / methodologies

1. ECNL / Data & Society: Recommendations for Incorporating Human Rights Into AI Impact Assessments https://datasociety.net/announcements/2021/11/22/recommendations-for-incorporating-human-rights-into-ai-impact-assessments/

2. Mantelero / Esposito: An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829759

3. Alan Turing Institute draft model for Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems - https://rm.coe.int/huderaf-coe-final-1-2752-6741-5300-v-1/1680a3f688

4. Netherlands, Fundamental Rights Impact Assessment:
https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms

5. Ada Lovelace Institute: Algorithmic impact assessment: a case study in healthcare
https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/

6. Microsoft Impact Assessment template: https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf

7. Meta:
https://www.bsr.org/en/our-insights/report-view/metas-expansion-end-to-end-encryption