



*Diversify with Intersectionally
Fairer Artificial Intelligence*

DIVERSIFIAR Policy Workshop

Ilna Georgieva, TNO Vector



Co-funded by
the European Union

AGENDA

Building the foundations

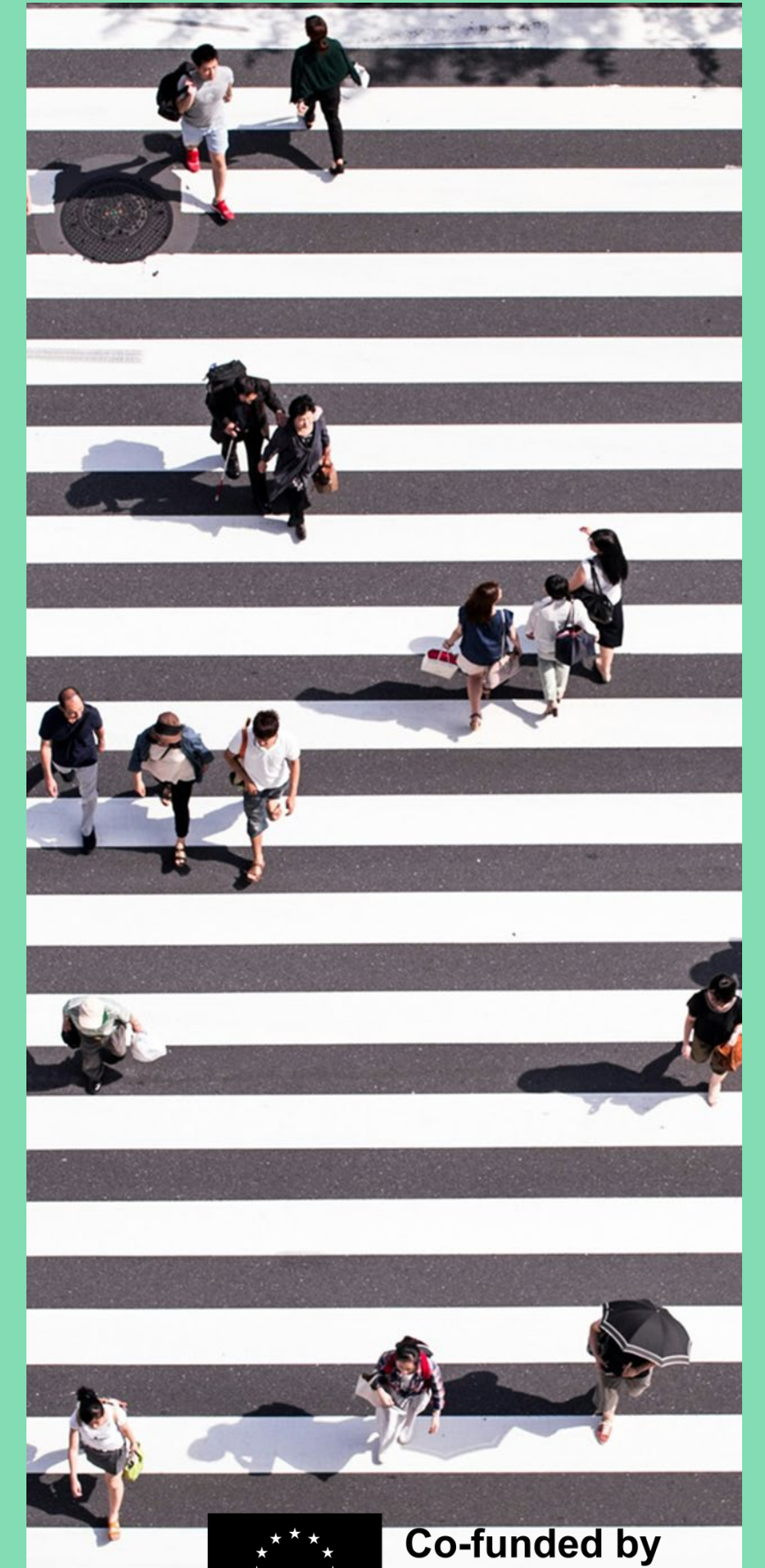
- Baseline knowledge on fairness in AI policy
- Risk-based governance and its limitations
- The need for understanding AI harm

Spotting the blind spots

- Why and how intersectional harms are often missed in policy and governance interventions
- Where do regulatory frameworks fall short

But first things first ...

What does
fairness mean
to you?
How do you
understand it in
relation to AI?



FAIRNESS IN AI GOVERNANCE

WHERE AND HOW IS AI FAIRNESS RELEVANT?



- Metrics, evaluations, bias detection and mitigation
- A very team-, data science- oriented effort
- Assumptions of causality
- Part of the first wave of AI governance
- A move towards fairness by design



SCANDAL

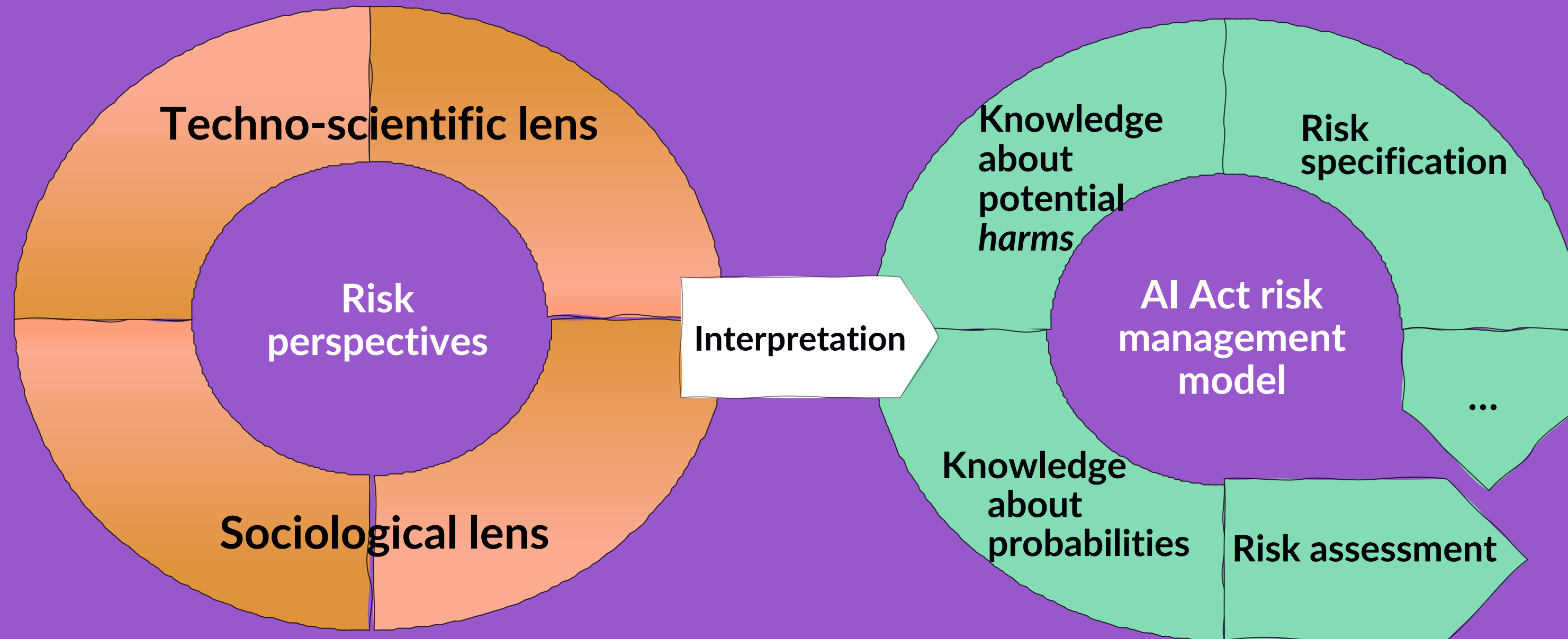
- Australia's Benefits Scandal Robodebt
- Dutch Child Benefits Scandal

TOWARDS A RISK BASED GOVERNANCE AND REGULATION OF AI

- An attempt at a coordinated approach in handling AI systems and their impacts
- The AI Act as a first direct, horizontal regulatory measure
- Effort in establishing risk classification systems – risk identification, assessment and management – and standardization (where risks incl. to fairness and discrimination play a role)

**“BUT THEN
WHAT DO AI
RISKS MEAN?”**

THE RISK MODEL OF THE AI ACT



GROUP REFLECTION: What harms are (in)visible in your own policy context?



THE NEED FOR NEW INSIGHTS AND NEW QUESTION

- Harms occur at the interplay of technical system components with societal power dynamics (Green & Viljoen, 2020), and travel through social systems.
- There is a need for identifying, comprehending and minimizing sociotechnical harms.

LET'S TALK ABOUT INTERSECTIONALITY FOR A MOMENT

OPERATIONALIZING INTERSECTIONALITY

What it facilitates

- A holistic lens, a framework for understanding overlapping identities
- Describes people's relational history in societal systems
- Traces the interactions between different sources of disadvantages or inequality
- Provides equal intervention opportunities -> informs the way to practice

What it does not

- Does not simply mean the added value of disadvantages
- Does not by default solve inequality and discrimination issues
- Does not provide answers but the questions on how to include marginalized people and counter their erasure

THE VALUE OF A COMPREHENSIVE AI HARMS TYPOLOGY

Clarifies the relationship between AI risk and harm, as current AI regulation has a heavy focus on risk management. AI harms knowledge can further help in litigation.

- Following the framing of the AI Act, defining risks and underlying harms will remain challenging regarding fundamental rights and values due to their inherent normative ambiguity.
- Tackles difficult normative decisions regarding accountability models and the so-called 'augmented harms'.
- It allows to explore other, more inclusive, governance pathways that are not effectively covered under the AI Act.



LIMITATIONS OF EXISTING REGULATION THROUGH AN INTERSECTIONAL LENS

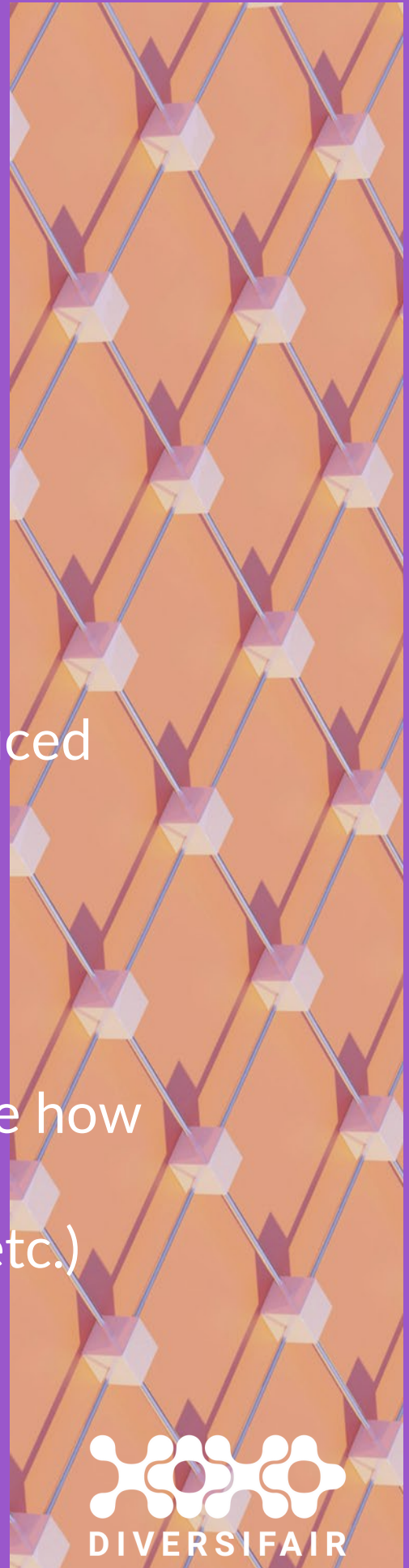
INTERSECTIONAL CRITIQUE OF THE GDPR AND THE AI ACT I

- **INDIVIDUALIZED RIGHTS VS. STRUCTURAL HARMS**

- The GDPR and the AI Act primarily focus on individual rights (e.g., consent, access, explanation), not on collective or systemic harms.
- Do not address group-based discrimination or historical inequalities that are reproduced through data and AI systems.

- **ONE-SIZE-FITS-ALL FRAMEWORKS**

- Both regulations use universal categories (e.g., “data subject”, “AI user”), which ignore how different communities are vulnerable to AI’s impact and harm.
- Vulnerable groups (e.g., migrants, people with disabilities, LGBTQAI+ communities, etc.) often lack the access, resources or literacy to claim their interests and rights.



INTERSECTIONAL CRITIQUE OF THE GDPR AND THE AI ACT II

- **TECHNO-CENTRICITY AND NARROW CONCEPTION OF DISCRIMINATION**

- Both regulations focus on explicit discrimination (e.g., treating someone differently based on protected characteristics), but often fail to capture indirect, cumulative or intersectional discrimination
- The AI Act emphasizes technical compliance without integrating social and historical contexts of discrimination; GDPR's reliance on automated processing clauses or legal bases often sidesteps real-world power imbalances in data use.

- **LACK OF INCLUSIVE PARTICIPATION**

- Marginalized communities are rarely meaningfully included in the development, design, deployment or governance of AI or data protection.
- This reinforces the priorities of dominant groups (e.g., tech companies, policymakers, the Global North), not those most affected.

But if this is all true, what are we missing on a structural basis? And how do we ask better questions?

THREE CHARACTERISTICS OF AI HARM

These characteristics can help researchers and policy makers to identify and address the structurality, systemic or procedural elements, that facilitate the manifestation of AI harms

Find the reading here

“Situating harms in the AI governance landscape—Understanding AI harms through the lens of interactivity, temporality and intentionality”



INTENTIONALITY

Relates to the nature of the AI harm and the quality of the mental state from which it originated: whether it was intentional, unintentional or the result of negligence

TEMPORALITY

Relates to the role of time in the way that AI harms come into being, e.g. singular or cumulative

INTERACTIVITY

Whether harm is the result of direct or indirect interaction with AI systems by the people that suffer the harm

FROM AI FAIRNESS TO PARTICIPATORY PRACTICES





*Diversify with Intersectionally
Fairer Artificial Intelligence*

GET IN TOUCH



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Culture Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



**Co-funded by
the European Union**