MAPPING HUMAN OVERSIGHT IN AUTOMATED DECISION-MAKING

A cross-jurisdictional taxonomy of regulatory roles: Netherlands, Australia and the European Union

THE UNIVERSITY OF MELBOURNE

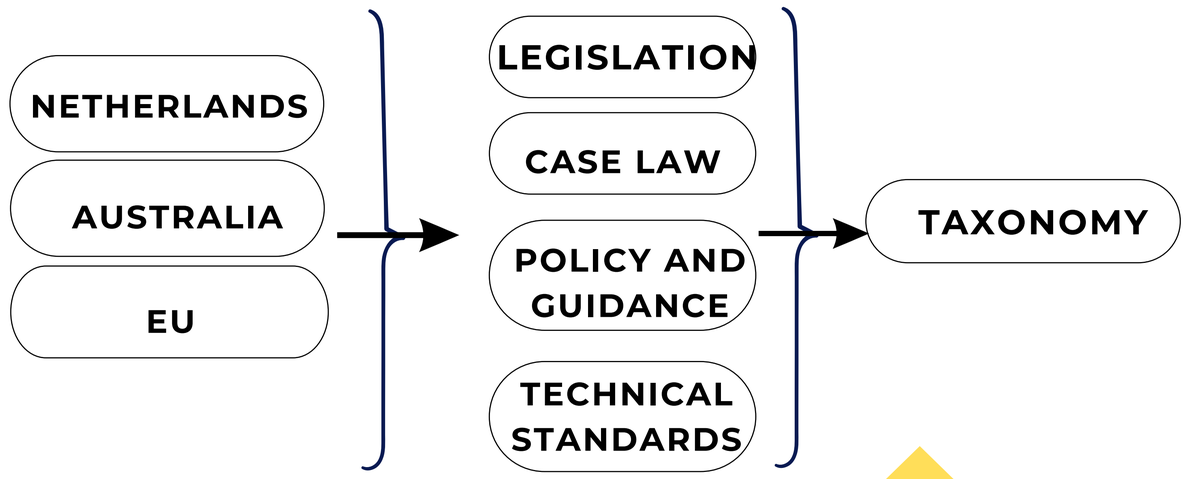
EMMA FINLAY

MELBOURNE LAW SCHOOL

INTRODUCTION

INTRODUCTION	Human oversight is widely involved in ADM regulation yet remains subject to varied definitions, interpretations, and expectations across jurisdictions and governance frameworks.
OBJECTIVE	Provide organisational clarification into human oversight's roles and highlight how the default anthropocentrism of ADM governance responses is fraught with inconsistencies across frameworks.
METHODOLOGY	An empirical, cross-jurisdictional analysis that maps how exactly human oversight is conceptualised and operationalised across the ADM governance lifecycle.

APPROACH



Taxonomy of Human Oversight's Regulatory Roles

19+ HUMAN OVERSIGHT ROLES

Example: Article 14.4 EU AIA: "...natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:
(a) to **properly understand the relevant capacities and limitations of the high-risk AI system** and be able to duly **monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;**
(b) to **remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system** (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
(c) to **correctly interpret the high-risk AI system's output,** taking into account, for example, the interpretation tools and methods available"

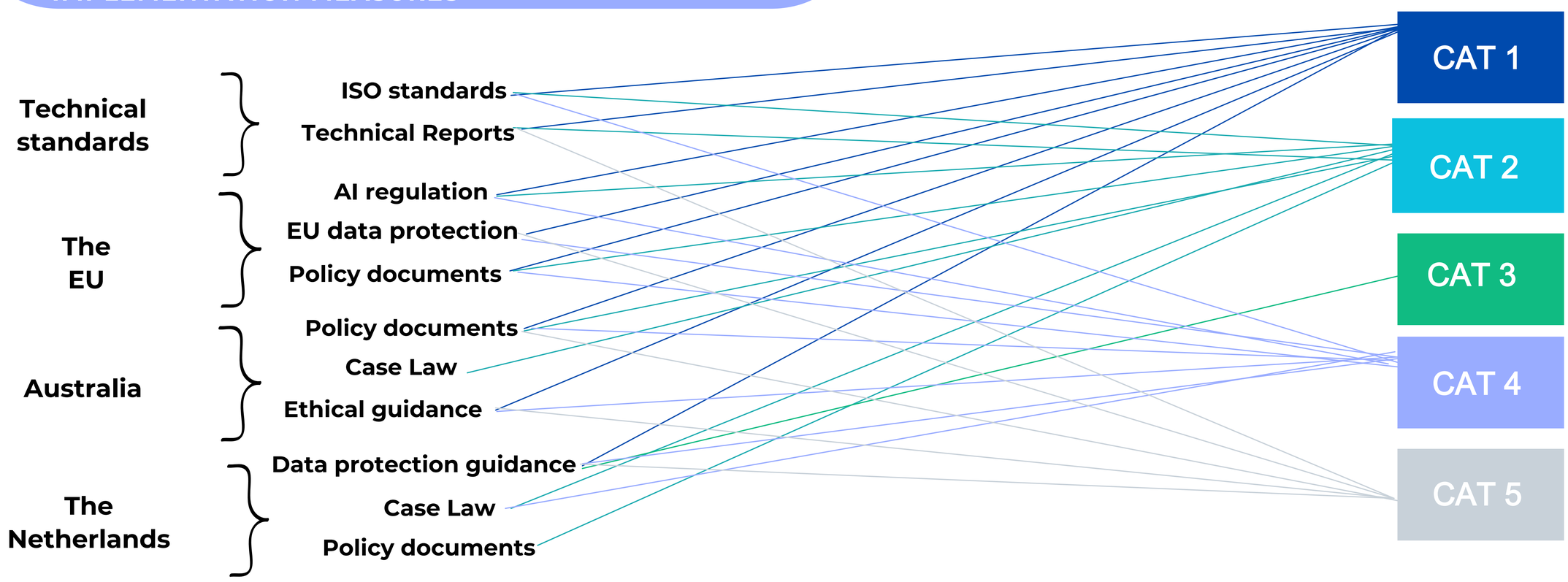
CATEGORY 1: EX-ANTE A DECISION
- OVERSEER
- MONITOR / SUPERVISOR
- INTERPRETER/MEANING MAKER
- REPORTER / FEEDBACK PROVIDER
- CONTROLLER

CATEGORY 2: DURING DECISION-MAKING PROCESSES
- ACTIVE DECISION-MAKER
- FINAL DECISION-MAKER
- INTERVENER/EX-POST REVIEW

CATEGORY 3: EX-POST A DECISION
- APPEAL ROLE

CATEGORY 4: PURPOSES, COMPETENTCIES, RESPONSIBILITIES
- EFFECTIVENESS
- UNDERSTANDING / AWARENESS
- ACCOUNTABILITY
- COMPETENCY / AUTHORITY / FITTING INTENTIONS
- PROTECT HUMAN DIGNITY
- ACCEPTABLE USE
- IMPLEMENTATION MEASURES

CATEGORY 5: INDIVIDUAL TO ORGANISATIONAL OVERSIGHT
- ORGANISATIONALLY DEFINED ROLE
- INDIVIDUAL TO ORGANISATIONAL OVERSIGHT
- INDEPENDENT OVERSIGHT BODIES



This research was conducted by the ARC Centre of Excellence for Automated Decision -Making and Society (CE200100005) and funded fully by the Australian Government through the ARC.