

# Legal Regulation of AI and Automated Decision-making in the Public Sector

Presentation for ANU SIS Symposium

15 October 2024

Dr Clement Chen

# Outline

- ‘General & hard’ laws governing the use of AI in the public sector
- Data Protection Law
  - Right not to be subject to automated decisions
  - Right to meaningful information
- Administrative Law
  - Right to Explainability – Duty to give reasons
  - Unfettered discretion – Limited role of automation

# Privacy & Data Protection Law

1. Regulating **personal** data for protecting (informational) privacy
2. Centralising *consent* in data collection and processing
  - But consent may be overburdensome for data subjects and outweighed by public interest
3. Data protection principles (DPPs) being developed for ensuring **fair** use of data in the AI context
  - EU's GDPR (General Data Protection Regulation): a light version of the rights *not to be subject to automated decisions* and *to human intervention*

# Australian Law Reform Proposals: Prohibition

## Privacy Act Review – Discussion Paper 2021

**17.1** Require privacy policies to include information on whether personal information will be used in ADM which has a legal, or similarly significant effect on people's rights.

### Question

- Should the concept of a decision with 'legal or similarly significant effect' be supplemented with a list of non-exhaustive examples that may meet this threshold?

## Attorney-General Report 2023

**19.1** Privacy policies should set out the types of personal information that will be used in substantially automated decisions which have a legal, or similarly significant effect on an individual's rights.

**19.2** High-level indicators of the types of decisions with a legal or similarly significant effect on an individual's rights should be included in the Act. This should be supplemented by OAIC Guidance.

**19.3** Introduce a right for individuals to request meaningful information about how substantially automated decisions with legal or similarly significant effect are made. Entities will be required to include information in privacy policies about the use of personal information to make substantially automated decisions with legal or similarly significant effect.

## Privacy Act Review

### Discussion Paper

October 2021

# Australian Law Reform Proposals: Right to Request Meaningful Info

## Attorney-General Report 2023

19.1 **Privacy policies** should set out the **types of personal information** that will be used in substantially automated decisions which have a legal, or similarly significant effect on an individual's rights.

19.3 Introduce a **right for individuals to request** meaningful information about **how substantially automated decisions** with legal or similarly significant effect are made. Entities will be required to include information in privacy policies about the use of personal information to make substantially automated decisions with legal or similarly significant effect.

Privacy Act Review  
| Report 2022

# Duty to Give Reasons under Administrative Law

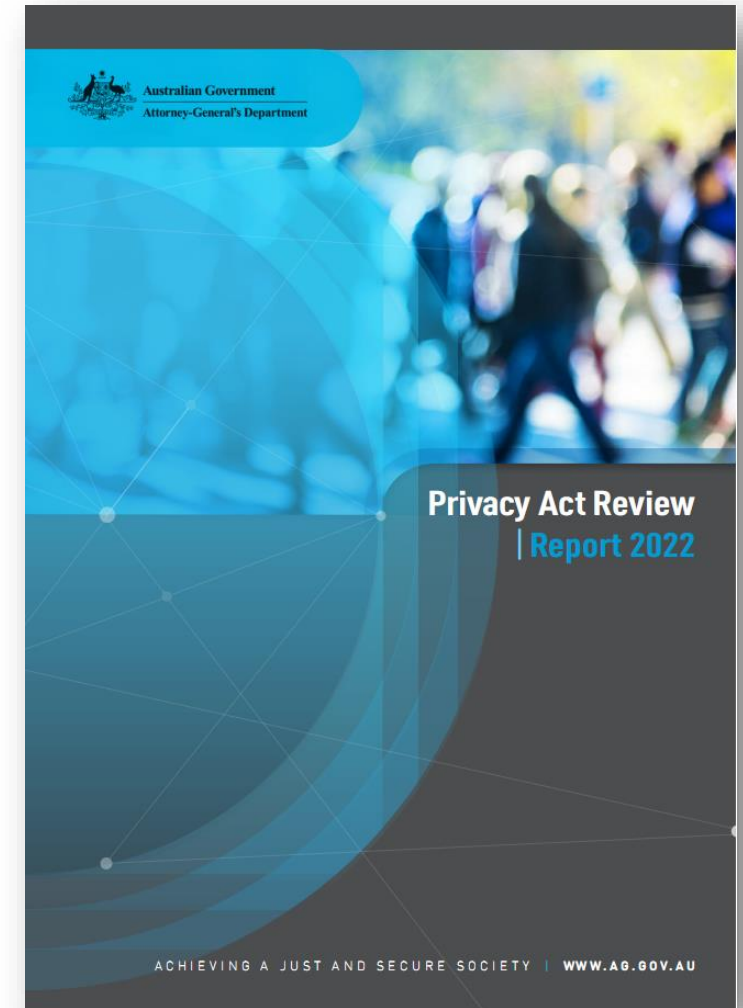
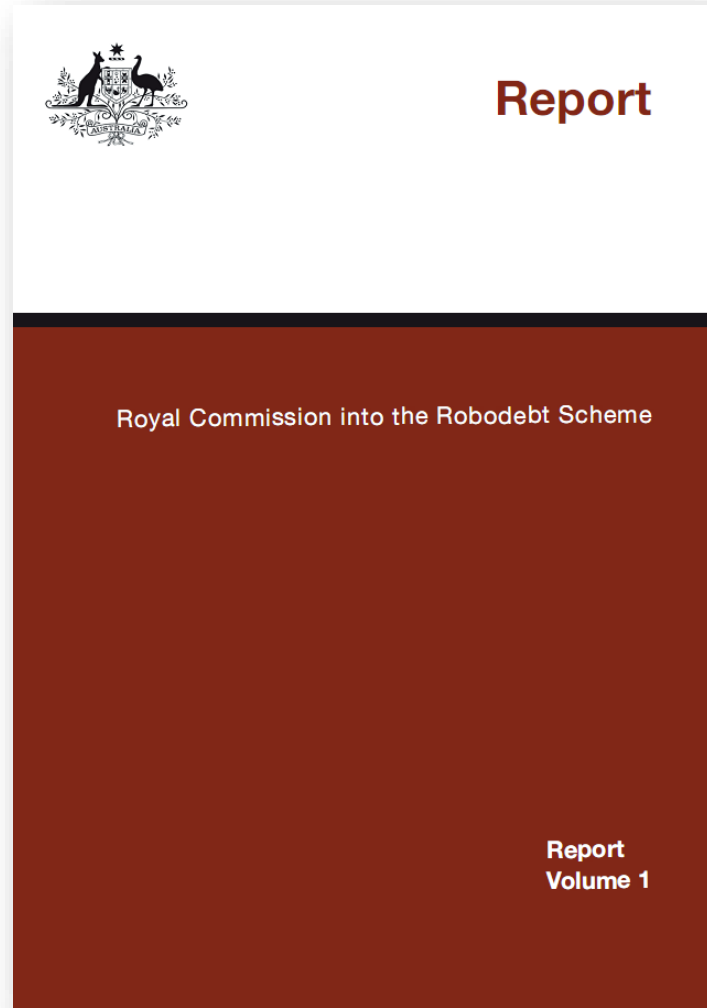
## Duty to give reasons by the decision-maker

- Natural justice in common law – the affected person’s right to be heard
- Statutory duty provided in civil law countries – ‘equality of arms’
  - Enabling the affected party to *understand* and *evaluate* the decision merits

## Quality of the reasons underlying an ADM

- Explanation of a particular decision v. overview of a complex model
  - Sensitive to specific conditions in which a statutory power is exercised
- Explanations of **how** v. **why** a decision was made

# An Australian Approach?



# AI Act 2024: Moving towards Specific AI Law in Europe

## Cross-sector regulation of AI

### Risk-based approach

- Harmful AI uses banned
- a defined list of “high risk” AI systems: subject to strict requirements (e.g. transparency)
- medium / low risk AI: lower-burden or no requirements for.

### Enforcement

- handled by a national regulators
- overseen by a new “EU AI Board”

Photo Source: Keepabl, <https://keepabl.com/news/infographic-eu-ai-act>

#### 4 RISK-BASED

'AI systems presenting only limited risk would be subject to very light transparency obligations, for example disclosing that the content was AI-generated so users can make informed decisions on further use.'



#### 5 HIGH-RISK AI SYSTEMS

Those with 'significant potential harm to health, safety, fundamental rights, environment, democracy and the rule of law' and those 'used to influence the outcome of elections and voter behaviour'. Mandatory fundamental rights impact assessment (FRIA).



#### 6 FOUNDATION MODELS / GPAI

Transparency requirements: technical documentation, detailed summaries on training content. Comply with EU copyright law. **High-impact GPAI models with 'systemic risk'**: conduct model evaluations, assess and mitigate systemic risks, conduct adversarial testing, ensure cybersecurity and report serious incidents to the EC, and report on energy efficiency.



#### 7 BANNED APPLICATIONS

Includes: biometric categorisation using sensitive characteristics; untargeted scraping of facial images for FR databases; emotion recognition in the workplace; certain social scoring; AI systems that manipulate human behaviour to circumvent free will or exploit people's vulnerabilities; and some cases of predictive policing.





# Regulatory Attempts Targeting Generative AI

European Union	The finalised AI Act <b>may include specific obligations for the providers of generative AI</b> and foundational models. Currently, the European Parliament position does not put these providers under the ‘high risk’ category, the triologue negotiations may give rise to changes.	General, EU(state)-led based
United States	<p>No specific work on generative AI. A recent bill in Congress on regulating AI has reached the committee stage. An executive order on responsible innovation of AI expected to be announced in the coming months.</p> <p>The Executive Order on Further Advancing Racial Equity and Support for Underserved Communities Through The Federal Government (16 Feb 2023) required that when designing, developing, acquiring, and using AI and automated systems in the <b>Federal Government, agencies shall do so ‘in a manner that advances equity’</b>.</p> <p>2022 White House’s Blueprint for an AI Bill of Rights.</p>	General, state-led
China	<p>Cyberspace Administration of China (CAC)’s Interim rules on Generative AI (effective 15 August 2023) stipulates that service providers shall ensure the generated content <b>adhere to the socialist core values</b> and not endanger national security, and shall take effective measures to <b>prevent discrimination based on ethnicity, belief, gender, age, professions, etc.</b> in the process of algorithm design, training data selection and model refinement.</p> <p>National Data Administration, a new national authority, was established on 25 Oct 2023.</p>	Quasi-general, state-led
Japan	<p>No plan (yet) for legislation or statutory rules on generative AI.</p> <p>Relying on the model of ‘agile digital governance’; <b>insulting AI model providers from copyright claims</b> under the existing Copyright Law.</p>	Sector-specific, industry-led

# AU Gov's Interim Response to Safe and Responsible AI



## Clarifying and strengthening laws to safeguard citizens

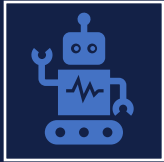
Significant work is underway or planned across the government to address issues raised during consultation on regulatory and policy frameworks. The department is working with agencies leading this work to ensure that views in submissions can inform these processes. This includes:

- developing new laws that will provide the Australian Communications and Media Authority with powers to combat online **misinformation and disinformation**
- an independent statutory review of the **Online Safety Act 2021** to ensure that the legislative framework remains responsive to online harms
- working with the state and territory governments, industry, and the research community to develop a regulatory framework for **automated vehicles** in Australia, including interactions with work health and safety laws
- ongoing research and consultation by the Attorney-General's Department and IP Australia, including through the AI Working Group of the IP Policy Group, on the implications of AI on copyright and broader IP law
- **implementing the privacy law reforms**
- strengthening Australia's competition and consumer laws to address issues posed by **digital platforms**
- agreeing an **Australian Framework for Generative AI in Schools** by education ministers to guide the responsible and ethical use of generative AI tools in ways that benefit students, schools and society while protecting privacy, security and safety
- ensuring the **security** of AI tools, such as using principles like security by design, through the government's work on the Cyber Security Strategy.

In addition, the Australian Government, as well as state and territory governments, will continue to consider areas where existing laws could be strengthened to address risks and harms posed by AI.

### Action

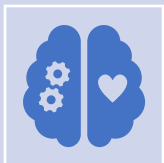
Building on recent (for example, privacy law) and proposed (for example, online safety and mis- and disinformation) reforms, the Australian Government will consider suggestions put forward in submissions on further opportunities to strengthen existing laws to address risks and harms from AI.



Right Not to Subject to a (Fully) Automated Decision  
[Data Protection Law]



Right to Explanation of Automated Decision-making  
[Administrative Law]



Unfettered Exercise of Discretion  
[Administrative Law]

# Importance of Existing Legal Frameworks

## Much attention paid to

- Legal gaps in some areas (e.g. deepfakes, plagiarism, IP of training data)
- ‘Soft laws’: voluntary principles, ethics and best practices guidelines
- Specific laws: (EU) AI Act 2024, (China) Administrative Rules of Gen AI 2023, (Canada) AI & Data Bill 2022

## But don't forget

- AI may not be a unique and standalone target of regulation
- The use of AI's different components can be subject to existing legal frameworks
- General (i.e. technology-neutral) laws about
  - Collection and use of personal data → data protection law
  - Automation of decision-making in the public sector → administrative law, human rights law...
  - Fair treatment of consumer → consumer protection law
  - Intellectual properties concerning training data and algorithmic models → copyright law, patents law...

## Higher Education Sector also subject to the **F**Acc**T** requirements

- **F**airness, **A**ccountability, **T**ransparency
- under data protection law and administrative law

# Accountability & Administrative Law

1. Accountability in administrative decision-making, including the use of personal/**non-personal** data, algorithms, and AI
2. Decision-makers shall provide:
  - justifications for its decision
  - reasonable exercise of discretion
  - corrections/remedies for harms caused to private parties
3. Administrative law has long-standing rules governing the public sector, including public universities and other research institutions
  - adaptive to the data-intensive environment

## Prohibition through a Right to Object?

GDPR Art. 22(1): Right not to be subject to automated individual decision-making

- '[t]he data subject shall have the right not to be subject to a decision based **solely on** automated processing, including profiling, which produces **legal effects** concerning him or her or **similarly significantly affects** him or her'

AUTOMATED INDIVIDUAL DECISION MAKING INCLUDING PROFILING		
ARTICLE	CONDITION	RELEVANT EXTRACT FROM PROVISION
Article 22(2)(a)	Contract	is necessary for entering into, or performance of, a contract between the data subject and a data controller;
Article 22(2)(b)	Authorised by law	is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
Article 22(2)(c)	Explicit consent	is based on the data subject's explicit consent.

# A Right to 'Human in the Loop' as a Substitute?

## A right to human intervention under GDPR Art. 22(3)

- For the cases of contract- or consent-based exceptions, 'the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the **right to obtain human intervention** on the part of the controller, **to express** his or her point of view and **to contest the decision.**'

Relate closely to the 'right to explanation'

# Schufa Case (C-634/21)

European Court of Justice December 2023



**SCORING & DATA**  
**AT SCHUFA**

Scoring explained - go the [Score Simulator](#) here

The seven most important score factors - simulated in seven steps. That's the Score Simulator.



# Schufa Case (C-634/21)

## Facts:

- OQ
  - applied for a loan but was rejected by the bank because her SCHUFA score was too low
  - requested to grant her access to the corresponding data
- SCHUFA
  - disclosed only the score and general principles
  - SCHUFA was not the decision maker; refusal based on trade secret grounds and the fear of gaming behaviour

## Held:

### 1. Scoring is a decision

- a) Produced automatically
- b) Established a probability value based on personal data concerning a person's ability to repay a loan in the future → to evaluate and predict → profiling (Art. 4)
- c) Draws strongly on by a third party(bank) → producing legal effects (Recital 71)

### 2. Right to obtain meaningful information under Art. 15

- a) Lack of protection of rights of the data subjects: entitled to receive 'meaningful information about the logic of the automated decision'
- b) Must provide suitable safeguards and to ensure fair and transparent processing

# An Emerging 'Right to Explanation' in Europe?

## GDPR

- Art. 12: Transparent information for the exercise of data subjects' rights
  - Tell data subjects about the existence of automated decision-making, including profiling.
- Art. 13: Notification obligation concerning the collection of personal data
- Art. 15: Right of access by the data subject
  - Not only the fact that profiling will occur but also **meaningful information about the logic involved in the ADM and the envisaged consequences** for the data subjects.
- (Non-binding) Recital 71
  - specifies that safeguards for data subjects 'should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, [and] **to obtain an explanation of the decision reached after such assessment and to challenge the decision**'

## Convention 108 + (Council of Europe)

- Modernised Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data
- Art. 9(1)(c): Every individual has a right to obtain, upon request, knowledge of the **reasoning underlying the data processing** where the results of such processing are applied to him or her.

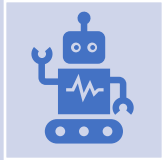
# Can Australian Privacy Principles Imply a Similar Right?

## APP 1 - Open and transparent management of personal information

- An entity should manage personal information in an open and transparent way. This includes having a clearly expressed and up to date [privacy policy](#).

## APP 12 – Access to Personal Information

- If an entity holds [personal information](#), it must, on request by the individual, give access to the information.
- This rule does not apply if the entity is
  - a (government) agency and is required or authorised to refuse access under Freedom of Information Act 1982 (Cth) or other Cth Acts, or
  - a (private sector) organisation and an exception applies such as:
    - The request for access is frivolous or vexatious
    - Giving access would have an unreasonable impact on the privacy of others
    - Giving access would reveal evaluative information generated within the entity in connection with a commercially sensitive decision-making process



Right Not to Subject to a (Fully) Automated Decision  
[Data Protection Law]



Right to Explanation of Automated Decision-making  
[Administrative Law]



Unfettered Exercise of Discretion  
[Administrative Law]

  
RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*

 **parcoursup**  
Entrez dans l'enseignement supérieur

 **EN COURS**

Du 01 juin au 7 juillet

**Phase d'admission**

*Parcoursup* adjudicated by France's

- Supreme Administrative Court (Conseil d'État) and
- Constitutional Council

# The *Parcoursup* Case in France

## Parcoursup: a digital platform

- created by the Law n°2018-166 to pre-register students in the higher education institutions.
- The 'local algorithms' are not publicly available as the platform allows institutions to introduce their own selection criteria.

## Legal requirements

- Under Art. L. 311-1 of the Code on Relations between the Public and the Administration (CRPA), as amended by the Law [for] a Digital Republic: the administrative authorities [concerned] are required to publish online or to communicate the administrative documents that they hold to persons who request them, under the conditions set out in this Code.
  - Also, Decree of 19 March 2019 requires the institutions to release the general criteria used in their selection process.
- But Art. L. 612-3 of the Code of Education provides that the right to obtain information regarding the criteria, procedures and pedagogical reasons applied to a final decision is reserved for the applicants concerned.

A student union sued a university to obtain the algorithm and the source code of Parcoursup

The first instance court (Administrative Court of Guadeloupe) held that such information should be disclosed to the student union.

On appeal, the highest administrative court (Conseil d'État) ruled that the CRPA provision was not applicable

- [Conseil d'État, 12 juin 2019, n°427916](#)

# Parcoursup – Cont'd

## Decision no.2020-834 of the Constitutional Council

### Issue

- (preliminary ruling requested by Conseil d'Etat on) the constitutionality of Art. L. 612-3 of the Code of Education about the limitation on the access to information

### Ruling

- Enshrines the right to communication of administrative documents as a constitutional right based on Article 15 of the Declaration of the Rights of Man and of the Citizen.
- Legislation may limit this right if justified by the general interest and the limitations are not disproportionate.
  - Limitation by the Code concerned to communication is justified by the secrecy of the deliberations
  - It is not disproportionate because guarantees are provided
- 'Interpretative reservation':
  - Must not to prevent third party's access to the criteria used for reviewing the applications once the national pre-registration procedure is completed.
  - Must specify the extent to which algorithmic processing was used to carry out this examination and respect the privacy of applicants.

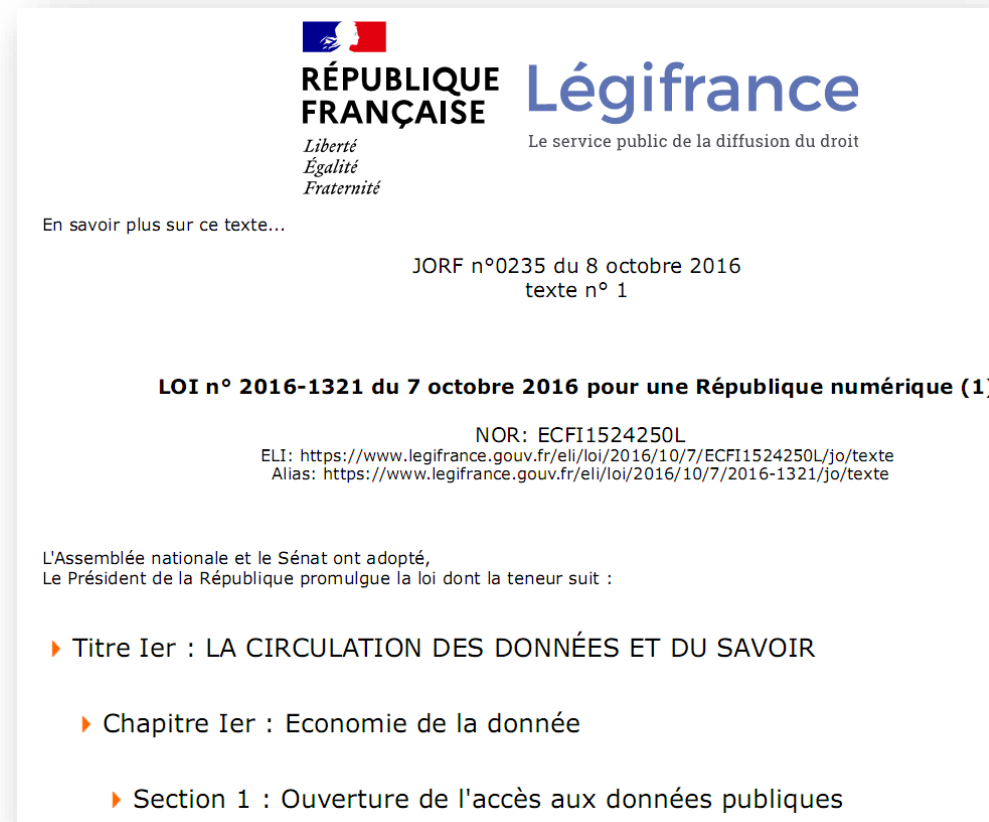
# Statutory Requirements of Explainability

## France

- Law No. 2016-1321 for a Digital Republic: right to explanation for administrative decisions taken on the basis of an algorithmic treatment
  1. the degree and the mode of **contribution of the algorithmic processing** to the decision making;
  2. the data processed and its source;
  3. the treatment parameters and, where appropriate, their **weighting**, applied to the situation of the person concerned;
  4. the operations carried out by the treatment
- 2016 Code Relations between the Public and the Administration
  - Art. L211-5: The statement of reasons [required] must be in writing and include a statement of the legal and factual considerations on which the decision is based.
  - Art. L311-3-1: An individual decision taken on the basis of algorithmic processing shall include an explicit statement informing the person concerned. The rules defining this processing as well as the main characteristics of its implementation shall be communicated by the administration to the person concerned if he/she so requests.

## Germany: VwVfG (Administrative Procedure Law)

- Every written (or electronic) decision requires an explanation or a 'statement of grounds' that outlines the essential factual and legal reasons giving rise to the decision.



The image shows a screenshot of the French legislative website, Légifrance. At the top, there is the French Republic logo and the text 'RÉPUBLIQUE FRANÇAISE' and 'Légifrance'. Below this, it says 'Le service public de la diffusion du droit'. The main content area displays the text of Law No. 2016-1321, dated October 7, 2016, for a Digital Republic. It includes the title 'LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)', the NOR number 'ECFI1524250L', and the URL 'https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte'. It also mentions that the law was adopted by the National Assembly and the Senate, and promulgated by the President of the Republic. The table of contents shows the title 'LA CIRCULATION DES DONNÉES ET DU SAVOIR', chapter 'Economie de la donnée', and section 'Ouverture de l'accès aux données publiques'.

See Edwards L and Veale M, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security Privacy 46; Olsen, Slosser and Hildebrandt, "What's in the Box?: The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration" in Oreste Pollicino et al (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021) 219-235.



# Turnitin Detection of AI-Generated Content

The screenshot displays the Turnitin AI Writing Report interface. At the top left, a blue banner reads "NOT FINAL - Work In Progress". The main title is "AI Writing Report". On the right, the Turnitin logo is visible. The central part of the report shows a document preview with highlighted text. To the right of the preview, a summary box states "56% detected as AI" and provides a "Submission Breakdown" bar chart. Below this, two categories of AI-generated content are listed: "1 AI-generated only" (24%) and "2 AI-generated text that was AI-paraphrased" (32%). Further down, "2 Signals of AI" are shown, including "Common AI phrasing" and "Unusual phrasing". At the bottom, there are links for "Learn more" and "Scroll for details".

**56% detected as AI**  
Percentage indicates the amount of qualifying text within the submission that was likely generated and/or paraphrased using AI.

**Submission Breakdown**

Category	Percentage
1 AI-generated only	24%
2 AI-generated text that was AI-paraphrased	32%

**2 Signals of AI (not factored into the AI Writing score)**

- 1 Common AI phrasing**  
This is a common phrase that Large Language Models use.  
"As an AI language model, I would like to point out..."
- 2 Unusual phrasing**  
This is a common unusual phrase that Large Language Models produce.  
"George W. Shrub" ↔ George W. Bush

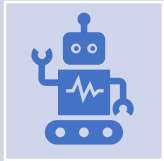
**AI Paraphrase detection**  
Deeper insight into how and where AI was used

**More evidence**  
Exploring additional signals that support AI detection

**Targeted guidance**  
Add more help, context and support with in-product guidance

© 2024 Turnitin LLC. All rights reserved.

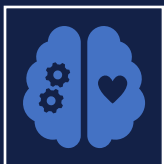
Scroll for details



Right Not to Subject to a (Fully) Automated Decision  
[Data Protection Law]



Right to Explanation of Automated Decision-making  
[Administrative Law]



Unfettered Exercise of Discretion  
[Administrative Law]

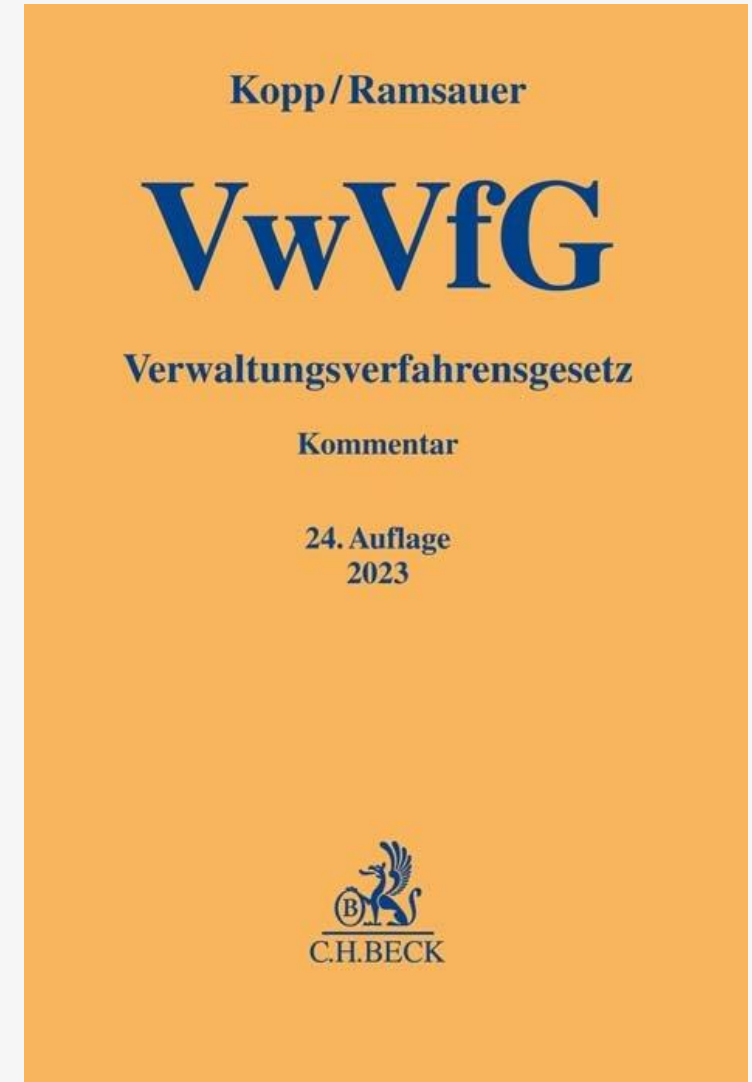
# Requirements of Unfettered Discretion

Discretion should be exercised and not to be fettered

- Case-by-case evaluation v. one-size-fits-all classification
- Exceptions from the general policy v. absolute adherence to instructed rules

Regulation of fully automated decision-making

- Approach 1: (legislative) ban
  - German VwVfG [Administrative Procedural Code] Art. 35a excludes the full ADM for cases involving discretion (since 2017)
- Approach 2: permission of ADM with no or minimal impact on individuals' rights
  - Poland: only positive visa decision
  - Latvia: only minimum fines provided by legislation can be imposed for an administrative offence recorded by technical means
  - Norway: ADM limited to decisions where little discretion is left to agencies; but not excluding delegation of powers to private actors if the latter do not have the normative power to prohibit or authorise activities.



# The *Pintarich* Case in Australia

*Pintarich v Deputy Commissioner of Taxation* [2018]  
FCAFC 79

## Facts

- An automated letter was sent to Pintarich concerning their tax debt owed, which was not checked by the tax official at the ATO.
- Pintarich paid the amount according to the letter and considered the tax settled. But the tax official claimed that the letter was incorrect in terms of failing to include the application of a general interest charge to settle a debt owed.

## Held

- A computer-generated letter declaring the amount and condition of remission of tax is not regarded as a 'decision' in legal sense,
- because it involves no 'mental process' of the official to whom the public power is entrusted.

ABC NEWS

Search...

Log In



Just In Watch Live Politics World Business Analysis Sport Science More ▾

## ATO executive admits letter automation error 'a bad look'

By business reporter [Nassim Khadem](#)

Posted Wed 24 Oct 2018 at 6:44pm



Australian Taxation Commissioner Chris Jordan faced questions at Senate Estimates. (ABC News: Mark Moore)

# Implications for the Higher Education Sector

Higher Education Sector should comply with the FAccT requirements under DP law and admin law

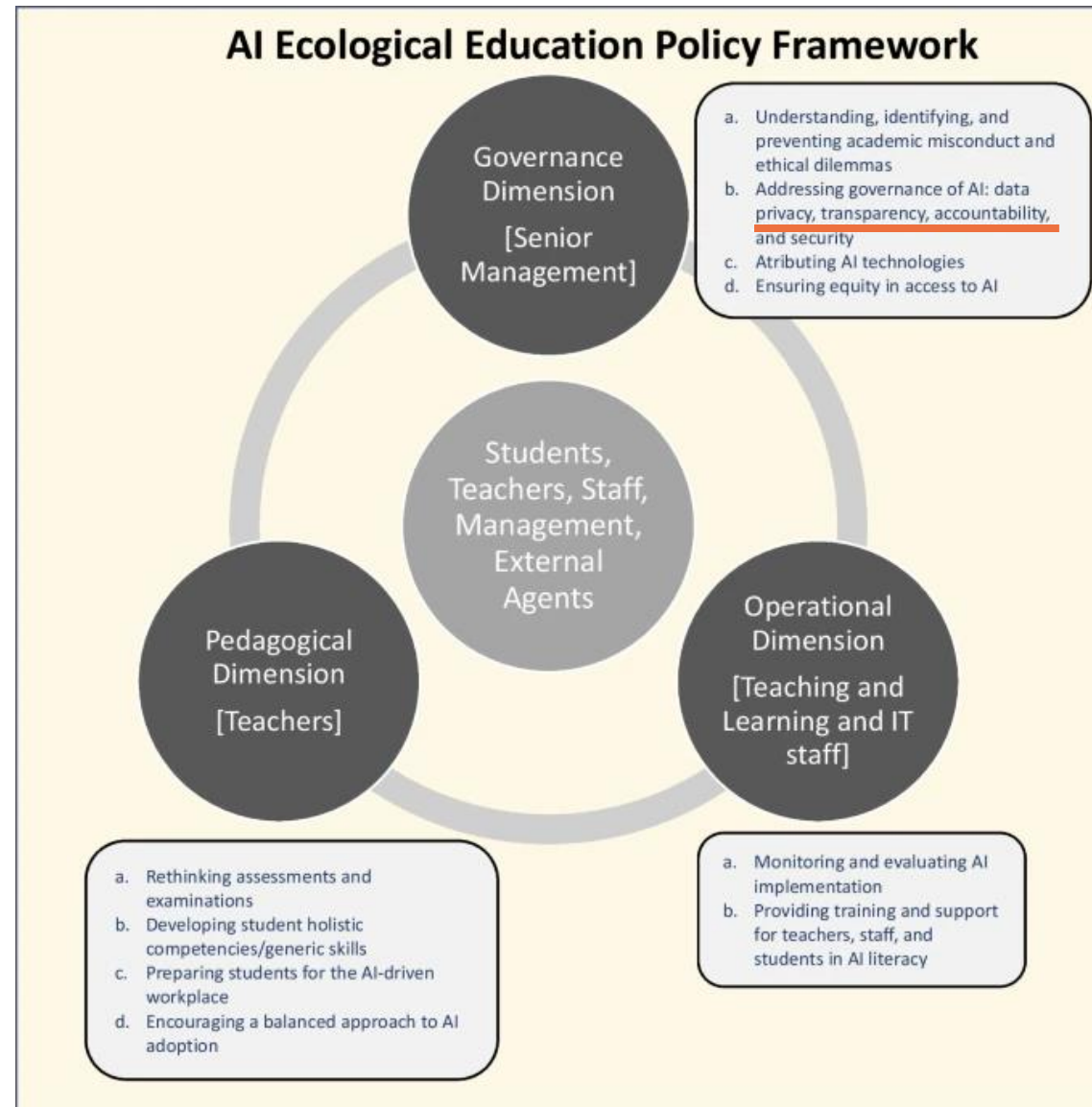
## Relevant scenarios

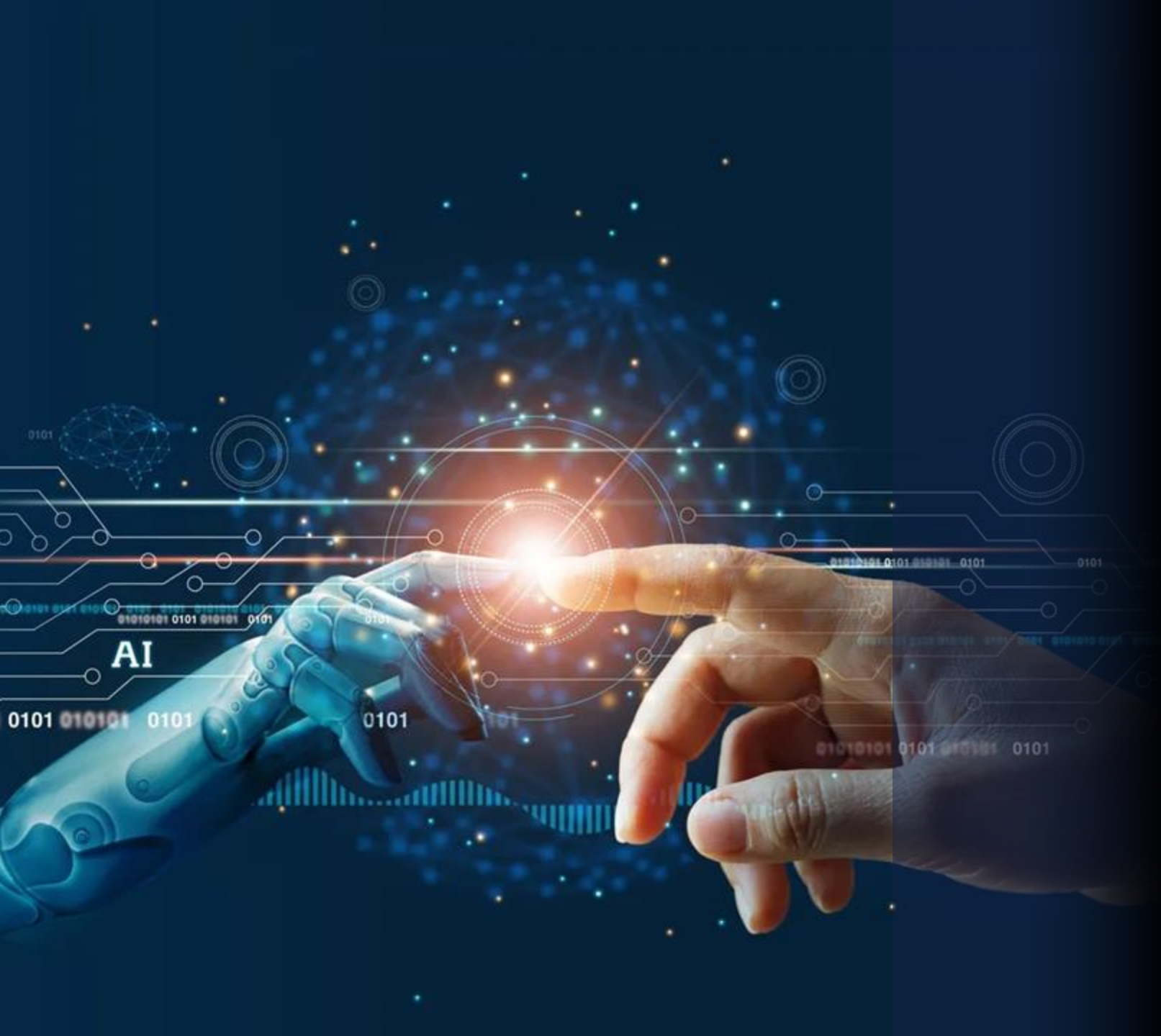
- Teaching and Assessment
  - ❑ Automated Admissions: right not to be subject to fully automated decision-making; challenges in contesting automated decisions
  - ❑ AI-Assisted Grading: unfettered discretion and duty to give reasons; challenges in identifying bias and explainability of AI decisions
  - ❑ Personalised Learning: privacy concerns; challenges in ensuring transparency in data usage
- Research and Student Support
  - ❑ Research Ethics (e.g. determination of plagiarism): compliance with data processing and decision-making requirements; challenges in providing clear methodological explanations
  - ❑ Predictive Analytics (e.g. of student behavioural patterns): consent to data collection and transparency in algorithms; challenges in the fairness of predictive models

## Developing institutional AI policies

- entails continuous understanding and evaluation of the legal landscape of FAccT issues
- whether general or specific

# Implications for the Higher Education Sector





# Discussion

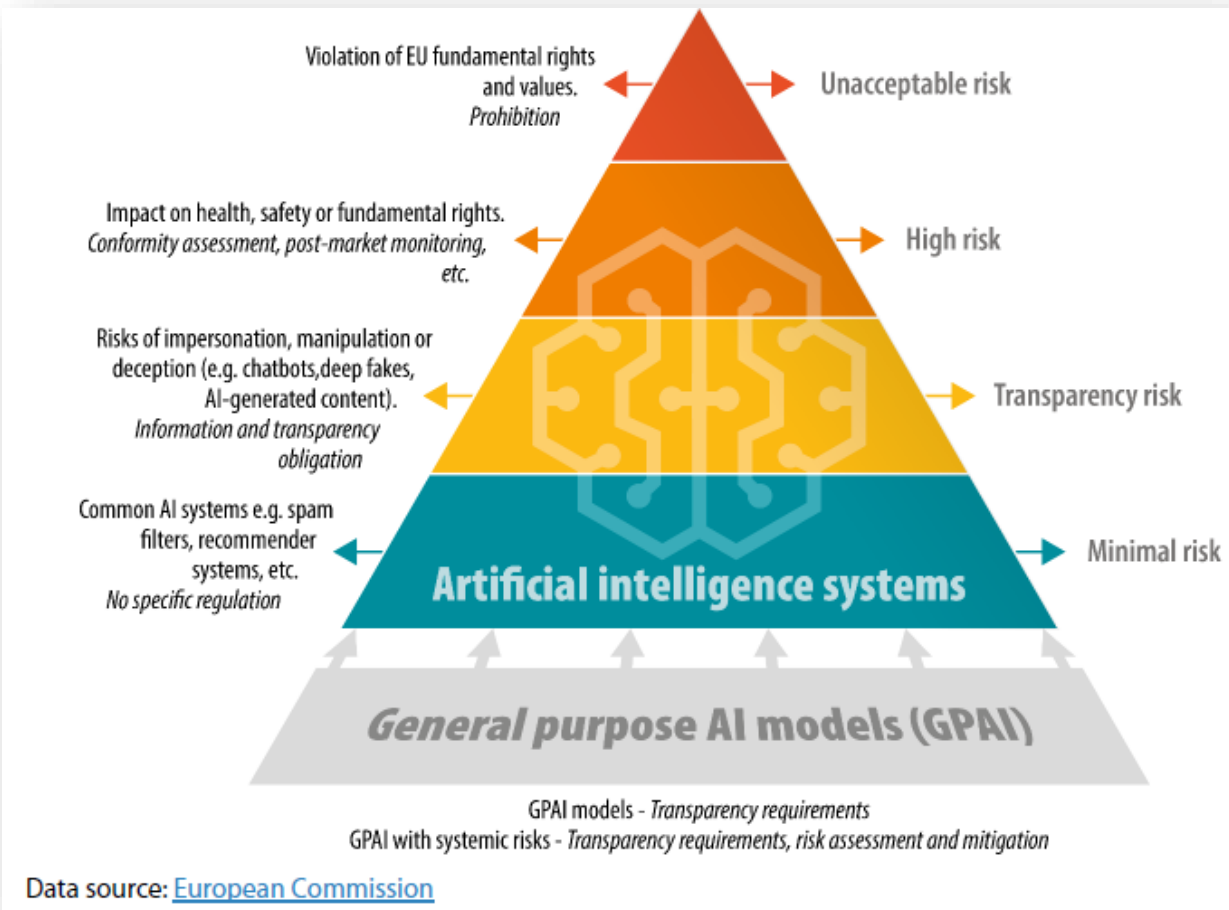
Thank you!

The image features three overlapping circles of a medium blue color, arranged horizontally. The circles overlap in the center, creating a darker blue area. The background is a dark gray. The word "Appendix" is written in white, sans-serif font, centered within the overlapping area of the circles.

Appendix



# Risk-based Regulation



	Public	Users	Documentation
identity; contact details			(assumed)
member states in use		(available publicly)	(available publicly)
purpose			
conformity assessment information			
relevant standards			
instructions for use			
human oversight & technologies			
accuracy, robustness, cybersecurity	"level of"; metrics (accuracy)		metrics; test logs; test reports
risky use circumstances			"detailed information"
performance on persons/groups			"detailed information"
input data		"where appropriate, specifications"	datasheets incl. training datasets and main characteristics; provenance; labelling procedures; data cleaning
pre-determined changes			"detailed description"; techniques to ensure "continuous compliance"
lifecycle information		expected lifetime; maintenance info	"description of any change made to the system"
post-market monitoring			"detailed description [of plan]"
risk management system			"detailed description"
design specifications			"general logic"; key choices and assumptions; optimisation function; trade-off decisions; description of hardware and interacting systems
methods and steps of development			role of pre-trained models/tools; computational resources used; training methodologies

Table 1: Main categories of information provided (or partially, or not) to the public, to users, and kept by providers in technical documentation. Not fully exhaustive and grouped for comparison; refer to the Act for full information.

These requirements are applied to "providers" as they must undergo *conformity assessment*.

# Identification of High-risk AI Systems

## Title III – High-risk AI Systems

### Article 6 Classification rules for high-risk AI systems

1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled:
  - a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II;
  - b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II.
2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.

# Identification of High-risk AI Systems – Cont'd

## Art 6 (2a): what are “high-risk AI systems”?

- AI systems will be considered high risk if they pose ‘a significant risk of harm, to the **health, safety, or fundamental rights** of natural persons’
- ‘Always’ considered high-risk – ‘performs **profiling of natural persons**’.
- If an AI system in Annex III is not assessed as high risk, providers are still required to register the product/service before it is placed on the market, and provide national competent authorities with documentation (if requested)

# Identification of High-risk AI Systems – Cont'd

## ANNEX III HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)

High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

1. Biometric identification and categorisation of natural persons:
  - (a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons;
2. Management and operation of critical infrastructure:
  - (a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.
3. Education and vocational training:
  - (a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions;
  - (b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.
4. Employment, workers management and access to self-employment:
  - (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
  - (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.
5. Access to and enjoyment of essential private services and public services and benefits:
  - (a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;
  - (b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;
  - (c) AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.
6. Law enforcement:
  - (a) AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;
  - (b) AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

- (c) AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3);
  - (d) AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences;
  - (e) AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups;
  - (f) AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences;
  - (g) AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.
7. Migration, asylum and border control management:
  - (a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;
  - (b) AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;
  - (c) AI systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features;
  - (d) AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.
8. Administration of justice and democratic processes:
  - (a) AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

# Stringent Transparency Requirements for High-risk AI Systems

## Title IV – Transparency Obligations For Certain AI Systems

### Article 52

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.
2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences.

# Banned Harmful AI Systems

## Title II – Prohibited AI Practices

### (I) Manipulative techniques

Article 5(I) The following artificial intelligence practices shall be prohibited:

- a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;
- b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;

# Fundamental Rights Impact Assessment under AI Act

Art 29(1) – scope of fundamental rights impact assessment covers high-risk AI systems as defined in Art 6(2)

- Prior to deploying,
  - bodies governed by public law or private operators providing public services,
  - and operators deploying high-risk systems – AI systems for assessing creditworthiness, risk assessment in life and health insurance
- Shall perform ‘an assessment of the impact on fundamental rights’ that use of that system may produce

# Data Protection Impact Assessment in the GDPR

## DPIA as a mechanism of accountability and additional protections

- Art. 35(1): 'likely to result in a high risk to the rights and freedoms of natural persons'; prior to the processing

## Consideration of 'fairness':

- In the determination of what is meant by 'high risk' in Art. 35(1) where a failure to conduct a DPIA properly (or indeed at all) would seemingly breach the fairness principle.
- In the interpretation of the situations identified as being specific cases of high risk in Art. 35(3).
- In the content of a DPIA including the items listed in Art. 35(7) and for example, the assessment of the proportionality and necessity of the processing operations, required under Art. 35(7)(b).

### Data protection impact assessment

1. Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.
2. The controller shall seek the advice of the data protection officer, where designated, when carrying out a data protection impact assessment.
3. A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:
  - (a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
  - (b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or
  - (c) a systematic monitoring of a publicly accessible area on a large scale.
4. The supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment pursuant to paragraph 1. The supervisory authority shall communicate those lists to the Board referred to in Article 68.
5. The supervisory authority may also establish and make public a list of the kind of processing operations for which no data protection impact assessment is required. The supervisory authority shall communicate those lists to the Board.
6. Prior to the adoption of the lists referred to in paragraphs 4 and 5, the competent supervisory authority shall apply the consistency mechanism referred to in Article 63 where such lists involve processing activities which are related to the offering of goods or services to data subjects or to the monitoring of their behaviour in several Member States, or may substantially affect the free movement of personal data within the Union.

7. The assessment shall contain at least:
  - (a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
  - (b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
  - (c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and
  - (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.
8. Compliance with approved codes of conduct referred to in Article 40 by the relevant controllers or processors shall be taken into due account in assessing the impact of the processing operations performed by such controllers or processors, in particular for the purposes of a data protection impact assessment.
9. Where appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations.
10. Where processing pursuant to point (c) or (e) of Article 6(1) has a legal basis in Union law or in the law of the Member State to which the controller is subject, that law regulates the specific processing operation or set of operations in question, and a data protection impact assessment has already been carried out as part of a general impact assessment in the context of the adoption of that legal basis, paragraphs 1 to 7 shall not apply unless Member States deem it to be necessary to carry out such an assessment prior to processing activities.
11. Where necessary, the controller shall carry out a review to assess if processing is performed in accordance with the data protection impact assessment at least when there is a change of the risk represented by processing operations.



# Privacy Impact Assessment in the Privacy Act 1988

s. 33D: Commissioner may direct an agency to give a privacy impact assessment.

- Agency means that this OAIC power is limited to the public sector.
- Needs to be 'significant impact on privacy'.
- Defines a PIA.

The screenshot shows the OAIC website page for the 'Guide to undertaking privacy impact assessments'. The page includes a navigation menu with links for 'About us', 'Privacy', 'Freedom of information', 'Information policy', and 'Consumer Data Right'. A search bar is located in the top right corner. The main content area features the title 'Guide to undertaking privacy impact assessments' and a date of '4 May 2020'. A featured image shows a hand holding a blue folder labeled 'Guide to undertaking Privacy Impact Assessments'. Below the image is a text block stating: 'This guide has a complementary e-learning course which aims to give you information on conducting a PIA in an easy-to-understand format so that you can have the confidence to do a PIA in your organisation or agency.' A yellow button labeled 'Launch the course' with a right arrow is positioned below the text. To the right of the main content is a 'Contents' section with a list of links: 'Introduction to privacy impact assessments', 'Undertaking a PIA', 'Glossary', 'Appendix A — Acknowledgments and resources', and 'Footnotes'. A 'Back to top' link is also present.

# Australian Law Reform Proposals: Impact Assessment

## Privacy Act Review – Discussion Paper 2021

### 11.1 – Option 1

APP entities that engage in the following restricted practices must take reasonable steps to identify privacy risks and implement measures to mitigate those risks:

- Direct marketing, including online targeted advertising on a large scale\*
- The collection, use or disclosure of sensitive information on a large scale
- The collection, use or disclosure of children’s personal information on a large scale
- The collection, use or disclosure of location data on a large scale
- The collection, use or disclosure of biometric or genetic data, including the use of facial recognition software
- The sale of personal information on a large scale
- The collection, use or disclosure of personal information for the purposes of influencing individuals’ behaviour or decisions on a large scale
- The collection, use or disclosure of personal information for the purposes of automated decision making with legal or significant effects, or
- Any collection, use or disclosure that is likely to result in a high privacy risk or risk of harm to an individual.

*\*‘Large scale’ test sourced from GDPR Article 35. Commissioner-issued guidance could provide further clarification on what is likely to constitute a ‘large scale’ for each type of personal information handling.*

## Attorney-General Report 2023



### 13.1 APP entities must conduct a privacy impact assessment for all activities with high privacy risks.

- A privacy impact assessment should be undertaken prior to the commencement of the high-risk activity.
- An entity should be required to produce a privacy impact assessment to the OAIC on request.
- The Privacy Act should provide that a high privacy risk activity is one that is ‘likely to have a significant impact on the privacy of individuals.’ OAIC guidance should be developed which articulates factors that that may indicate a high privacy risk, and provides examples of activities that will generally require a privacy impact assessment to be completed. Specific high-risk practices could also be set out in the Act.

# AI Act in EU's Evolving Regulatory Matrix about Data

Multiple legislation to establish an 'European way of data governance'  
anticipated in the EU Data Strategy (COM(2020) 66 final)

## EU laws proposed or enacted

AI Act Proposal (2021)

Data Act Proposal (2022)

- Right of access to and use of data generated by Internet of Things (IoT) devices
- 'data spaces' (general and sector-specific) as special arrangements for data sharing

Data Governance Act (effective Sep 2023)

- Re-use of data by public sector bodies
- Voluntary sharing of 'protected data', establishing mechanism of data 'altruism'
- re-use of public sector data also regulated by the Open Data Directive (effective 2019)

Digital Markets Act (effective May 2023)

- Regulating gate-keepers (i.e. leading internet service providers) and enhancing competition

Digital Services Act (effective Feb 2024)

- Regulating content and liability of intermediary service providers (against illegal content and protecting Internet users' rights)

# Legal Subjectivity Threatened by ADM

## Regulatory aims

### Epistemic threats

- Distortion of legally relevant characteristics/factors
  - ‘flattening’ of the high dimensional reality into machine-readable data points
  - Disregard of traits unique to the subject but not shared by the category of people whose profiling applies to the subject
- Including characteristics/factors not anticipated or comprehensible
  - Factors being statistically correlated but not argumentatively pertinent
  - Obscure pattern generated by machine



Right to correction



Right to explainability

### Agential threats

- Evaluation essentially as behavioural prediction based on existing data
  - Practically denying free will
  - Indifference to the subject’s state of mind (essential for attributing blame or confirming consent)
- Reducing the participatory opportunity
  - bring in new relevant factors
  - contest the evaluation



Right not to be subject to ADM



Right to non-discrimination, proportionality, etc



Right to hearing



Right to contest

# Granularity of the Explanations

	Contagion Prediction	Predictive Policing
Impact	Contemporarily restricts mobility	Arresting suspects
Complexity	Decision tree, based on <i>classifiers recommended</i> by epidemiological studies	Unsupervised machine learning, based on <i>Behavioural patterns</i> of the arrested people in given districts
Threshold of explainability	Medium	Heightened

How *granular* should the explanation be?

- Impact on individual's rights and interests
- Complexity of algorithms