22 March 2021

CIPL Recommendations on Adopting a Risk-Based Approach to Regulating Artificial Intelligence in the EU

The Centre for Information Policy Leadership (CIPL)[1] has been at the forefront of promoting organisational accountability and a risk-based approach as cornerstones of effective data protection law, policy and oversight for more than 20 years. With the fourth industrial revolution and accelerated digitalisation and datafication of our society and economy, these two concepts will play an increasingly important role **in all areas** of digital policy, law and regulation in the EU, especially regarding the use of artificial intelligence (AI), because:

- They are critical to building and delivering trust in the modern digital age; and
- They deliver a balanced and future-proof approach to regulation that enables private and public sector organisations in the EU to adopt the latest technologies and maximize their benefits in a responsible and human-centric way.

Building on its prior work,[2] CIPL has been working with experts in the EU and multinational companies who are leaders in AI to collect best practices and emerging trends in AI accountability. CIPL's objective is to inform the current EU discussions on the development of rules to regulate AI. This paper summarises CIPL's vision on how to implement a risk-based approach to AI regulation and compliance. It is based on the premise that AI regulation must remain agile, just like the technology uses that it seeks to regulate. Hence, it should not aim for a one-size-fits-all approach or elimination of all risks. It must allow for the evaluation of contextual risks and benefits, mitigation of risks, honest error and constant improvement.[3]

CIPL's vision for an effective and future-proof AI framework benefits:

- **Individuals,** by ensuring that the benefits of AI are continuously balanced against potential risks;
- **Organisations,** by enabling them to reap benefits from the latest technologies and stay competitive in the modern digital age without harming individuals and society and, where possible, providing benefits to these groups;
- **Regulators,** by providing modern tools that enable effective regulation without undermining innovation; and

[1] CIPL is a global data privacy and cybersecurity think tank in the law firm of Hunton Andrews Kurth LLP and is financially supported by the law firm and 80 member companies that are leaders in key sectors of the global economy. CIPL's mission is to engage in thought leadership and develop best practices that ensure both effective privacy protections and the responsible use of personal information in the modern information age. CIPL's work facilitates constructive engagement between business leaders, privacy and security professionals, regulators and policymakers around the world. For more information, please see CIPL's website. CIPL generally develops its white papers and public consultation responses with input from its member companies. Nothing in this submission should be construed as representing the views of any individual CIPL member company or of the law firm of Hunton Andrews Kurth.

[2] See CIPL's Response to the EU Commission White Paper - How to Leverage the GDPR, Accountability and Regulatory Innovation in AI Development, Deployment, and Uptake; How the General Data Protection Regulation (GDPR) Regulates AI. Artificial Intelligence and Data Protection in Tension, Hard Issues and Practical Solutions in AI

[3] The approach to regulating AI should be aligned with the Commission's approach to defining a framework for data access and use that calls for "an agile approach to governance that favours experimentation" over "heavy-handed ex ante regulation."

- **Society and the economy,** by fostering broader development and adoption of beneficial and trustworthy AI, taking into account the opportunity cost of not using AI.

In this paper, CIPL recommends a risk-based approach to regulating AI applications comprised of: (1) A regulatory framework focusing only on AI applications that are "high risk"; (2) A risk-based organisational accountability framework that calibrates AI requirements and compliance to the specific risks at hand; and (3) Smart and risk-based oversight.

This paper does not address the liability regime designed to provide rules for compensation to natural or legal persons if actual harm has occurred.

1. Determining what AI applications will be covered by the regulatory framework

Any AI regulation should focus only on high-risk AI applications and be proportionate and coherent in relation to the existing body of laws across the EU that apply in a technology-neutral manner. Therefore, there needs to be a way for organisations to determine whether their AI applications meet the threshold high-risk criteria for falling within the scope of the regulation. CIPL proposes that the regulation take the following approach to determining whether AI falls within or outside the scope of the regulation:

1.1 Adopt an easy-to-use framework for identifying covered high-risk AI applications. The approach for identifying covered "high-risk" AI applications must work for organisations of all sizes. It should not be too complex, prescriptive or multi-layered, such as the risk assessment model of the German Data Ethics Commission, which would be disproportionate for most organisations, difficult to apply in practice, and may hamper the development and deployment of innovative AI technologies in Europe.

1.2 The framework for identifying covered high-risk AI applications should involve the use of impact assessments designed to assess the likelihood, severity and scale of the impact of the AI use. Such impact assessments would include the following considerations:

- Severity and likelihood of harm to individuals, groups, or society at large (relying on conclusions that can be reached with reasonable certainty);
- Level and meaningfulness of human involvement and review and appropriateness given the context;
- Magnitude and likelihood of benefit of the AI use for individuals, groups, or society at large;
- Reticence risk and/or opportunity costs of not using the AI for individuals, groups of individuals, or society at large. This would include weighing of benefits of the AI use versus leaving the process under the current status quo (i.e., measuring whether the outcome is enhanced by the use of AI rather than leaving it as currently done); and
- Mitigation measures to address the risks.

1.3 The regulation should provide criteria and guardrails to organisations for determining high-risk AI applications. For example, it could:

- Specify categories, types and examples of harm to individuals and their fundamental rights that should be considered in the impact assessments, including physical, material or non-material damage. These could include discrimination, identity theft, financial loss, loss of

confidentiality in case of professional secrecy, unauthorised reversal of anonymisation or pseudonymisation;[4]

- Refer to criteria where AI is used to make a decision, based solely on automated processing, which produces a legal effect or similarly significant effect to a legal effect;[5]
- Refer to relevant criteria of high risk found in sectoral laws;
- Refer to criteria of high risk found in soft law instruments, such as the High Level Expert Group's Trustworthy AI self-assessment list; or
- Refer to relevant criteria found in co-regulatory tools, such as codes of conduct and certifications (including at the sectoral level).

1.4 **An "AI innovation board" should provide additional guidance and referentials to assist organisations in identifying high-risk AI.** This advisory board would gather the relevant authorities (including sector specific authorities), the EU Commission, industry (including SMEs and start-ups) and civil society representatives with the mandate to promote the development of trustworthy AI products and services across the EU.[6] The board's missions, overseen by the European Commission, should include developing, publishing and regularly updating lists of criteria for, and use cases of, high-risk AI or taxonomy of harms. Its role would also be to promote best practices for mitigating risks, ensuring EU-wide consistency in the approach to defining high-risk AI applications and promoting international interoperability. Organisations should also have the possibility to consult the board on specific use cases. The board would have to be set up and start working immediately after the regulation is adopted, and before it enters into force, so that it is fully operational at the time the regulation becomes effective.

1.5 **Illustrations of high-risk AI applications provided in the regulation or regulatory guidance should be treated as rebuttable presumptions**. This would enable organisations to take account of the highly contextual nature of AI applications and give them the opportunity to demonstrate that the use of an AI application in a specific context does not present a high risk. For example, using AI to assess diabetic retinopathy, as part of a triage process for initial screenings that reduces the risk of high priority patients having to wait weeks for ophthalmologists to review imagery, may not necessarily involve a high risk, as the AI application is intended to perform a triage and not to provide a final diagnosis. Conversely, relying solely on AI to diagnose diabetic retinopathy and instigate treatment, without any additional medical review, may be high-risk AI.

1.6 **Pre-screening or triage assessment.** Organisations would perform a simple pre-screening or triage assessment to determine whether a full-scale impact assessment is necessary in light of the criteria provided in the law and guidance. This would allow organisations to better allocate their resources to the assessment of AI applications that may carry a high risk and prevent organisations from undertaking assessments of AI use in contexts where it is obvious that there is very little risk involved.

1.7 **Impact assessment has been shown to be effective in AI product development for all sizes of organisations; it can also be effective in identifying "high-risk" AI applications.** The recent Open

[4] See, for example, Recital 75 GDPR.
[5] See Article 22 GDPR and Appendix 2.
[6] This board could take inspiration for instance from the existing ENISA in the field of cybersecurity.

Loop project confirmed the benefits of conducting impact assessments for European start-ups. This project gathered ten AI start-ups from diverse sectors to conduct AI impact assessments related to their AI products, based on principle-based regulation and procedural regulatory guidance. All of the participants were better able to identify the risks posed by their AI application based on practical regulatory guidance, mitigate these risks and embed best practices and safeguards in the design of their products. This resulted in greater efficiency and faster delivery to market, while reducing costs and risks of later disruption. This project demonstrates that impact assessments that are based on operational guidance provided through soft law can be an effective mechanism to identify high-risk AI applications for purposes of an EU AI regulation.[7]

1.8 **As part of the impact assessment, organisations using AI should factor in the benefits of an AI application versus the status quo.** In some instances, the benefits of an AI use to individuals, or a group of individuals, may be significant regardless of its risks. While the benefit of the AI use should not directly affect the risk classification of an AI application, consideration of the benefit would reduce the reticence risk of not going forward with the intended beneficial AI application merely due to the possibility of high risk. A balancing between benefits and risks could be performed similarly to the legitimate interest test of Article 6(f) GDPR. In the context of AI, this test would require an organisation to weigh the legitimate interests of using an AI technology (for the organisation, individuals, groups of individuals, society) against the interests or fundamental rights of individuals to ensure both benefits and risks are considered and weighed against each other in the development and implementation of a given AI application. Both the impact assessment and the legitimate interest-balancing test are well known and familiar practices under the GDPR for organisations in the EU. As such, it would be easy to leverage these existing practices and procedures for the purposes of assessing the risks and benefits of an AI application. These assessments would help identify a potentially compelling need to proceed with a high-risk AI application, would demonstrate how it can improve an existing (non-AI) process, and would help identify what safeguards might be devised to mitigate the high risk at hand. Such approach would allow European organisations of varying sizes to leverage existing tools and processes, rather than having to absorb new formalistic requirements.

1.9 **Make explicit that AI uses with no or low risk are outside the scope of the AI regulation.** Naturally, a regulation focusing on high-risk AI applications would not impact beneficial uses of AI that entail very low or no risk use. Examples of such no or low risk uses include AI used for the following purposes:

- For industrial or technical purposes, including streamlining existing processes, outside of high stakes settings;
- To ensure network security and prevent cyber-attacks;
- To prevent and detect fraudulent financial or commercial transactions;
- To prevent or detect unlawful money laundering;
- To anticipate the likelihood and nature of customer complaints to target appropriate proactive customer service;
- For website and device audience measurement purposes to ensure compliance with advertising standards (e.g., requirements not to advertise foods high in fat, sugar and sodium when the audience is comprised of more than 25 % children);

[7] https://ai.facebook.com/blog/introducing-open-loop-a-global-program-bridging-tech-and-policy-innovation/

- For website content management and moderation purposes (for identifying content that may be harmful or illegal);
- To provide for spam filtering;
- To improve search engines performance;
- To improve language understanding, speech recognition, image understanding, audio understanding, and hand writing recognition;
- To provide and improve machine translation;
- To optimise storage facilities management;
- To organise transport and logistics;
- To increase responsiveness in case of public calamity;
- To understand effects of climate change, optimise use of resources (e.g., electricity, compute power) to mitigate effects of climate change, and identify new solutions to address climate change (e.g., more efficient batteries, new solar cell materials); and
- To improve manufacturing management and predictive maintenance.

1.10 **Whether an AI use is considered high risk should not solely depend on the industry sector using the AI.** As AI technology evolves very rapidly and traditional sectors have become very dynamic and are less and less clearly delineated, these classifications will become very quickly outdated. For instance, an AI-based application in an autonomous vehicle that measures the behavior and health of a driver (temperature, focus, alertness), used by the automotive industry, could be used in other ways, such as by the health industry or public health authorities to fight a pandemic. In addition, AI innovative models are often built by SMEs and start-ups and are deployed at scale by larger organisations in a variety of use cases (for instance AI models for automated recognition of objects can be used for various industries from e-tail to security enhancing tools).

2. Risk-based organisational accountability to calibrate AI requirements and compliance

The following elements and considerations apply to those high-risk AI applications that have been established as within the scope of the regulation. They are not designed to define the requirements applicable to high-risk AI applications, but more to explain how these requirements should be defined in law and implemented in practice by organisations.

2.1 **Principle- and outcome-based rules.** Given the pace of AI evolution, to avoid the regulation becoming quickly outdated, the AI regulatory approach should avoid imposing prescriptive requirements. Instead, it should provide for principle- and outcome-based rules that enable organisations to progress towards the achievement of specified outcomes (e.g., fairness, transparency, accuracy, human oversight) through risk-based, concrete, demonstrable and verifiable internal measures.

2.2 **Provide obligations of means or process.** The AI regulation should focus on and enable effective processes that lead to the desired outcomes. Some of the outcomes (such as fairness, transparency, and accuracy) are subject to trade-offs in particular contexts, evolve over time, and may be challenging to reach and to maintain. Thus, rather than imposing concrete targets for specific metrics (which will be very hard to generalise appropriately), organisations should be encouraged and rewarded for reaching these desired outcomes or getting closer to them as much

as possible, through monitoring and ongoing improvement and adaptation of relevant mitigation measures. In addition, requiring an AI application to be fair and unbiased towards a certain group of individuals will require effective verification of the absence of bias based on sensitive data that may not be available or cannot be legally processed (as processing special categories of data under the GDPR is subject to significant restrictions).

2.3 Include an explicit accountability obligation. CIPL recommends the inclusion of a specific accountability requirement in the AI regulation as follows: *"Taking into account the nature, scope, context, purposes, impact, risks and benefits of an AI application, the organisation shall implement, and be able to demonstrate that it has implemented, appropriate organisational and technical policies and measures to appropriately mitigate the risks while enabling compliance with the principles in the AI Regulation. Organisations will review and update such policies and measures where necessary."* Including an explicit accountability requirement will result in organisations being more thoughtful about the risks and impacts of their AI applications and will help them establish processes and controls to develop and implement responsible, trustworthy and sustainable AI applications.[8]

2.4 Calibrating compliance with the legal requirements based on risk and benefits. Organisations must be able to calibrate the requirements of the regulation based on the outcome of the impact assessment. The higher the risk, the stronger and more sophisticated the implementation of a particular requirement and accountability measures and controls have to be. Organisations are best placed not only to assess their risks, but also to define, test, apply and verify the effectiveness of the controls and mitigating measures depending on context. Depending on the type of AI use, organisations may weigh risks differently and may have to decide on trade-offs between the different outcomes, mitigating measures and frequency of their reviews. Careful consideration is also needed in regards to any significant impact that mitigations could have on the likelihood and degree of benefits being realised.

2.5 An agile regulatory framework must allow for continuous improvement. AI technologies are rapidly evolving, and risks and benefits may be impossible to predict at the outset. Organisations have to monitor the performance of their AI application and regularly adapt, reiterate and improve it, fixing issues as they appear. Therefore, the regulatory framework should be flexible enough to permit this agility. It should encourage organisations to identify risks, address them and adapt their mitigation measures throughout the life cycle of an AI application in an iterative manner. The regulation should also allow for the possibility of further regulatory changes and "tweaks" at certain intervals, in consultation with the AI innovation board, industry bodies and stakeholders involved in developing and deploying AI technologies.

2.6 Limited prior regulatory gatekeeping. Prior consultation with regulators or prior conformity assessments should be limited to only high-risk AI uses where risks cannot be sufficiently mitigated and residual risks remain high (such as the use of facial recognition for unique identification purposes in public settings, or public sector uses of AI in policing). Limiting prior

[8] Many other compliance areas, such as anti-bribery, anti-money laundering, export control or medicine and food regulation already require organisations to implement comprehensive risk and compliance management programs. See CIPL paper Organisational Accountability – Past, Present and Future

consultation and assessments would avoid burdensome, inefficient and lengthy administrative procedures not suited for fast-paced development of AI applications and systems. It would also trigger responsible behaviors from organisations by encouraging them to provide more mitigation and would avoid deterring organisations from engaging in beneficial AI uses. As a reminder, the GDPR moved away from ex ante notifications to regulators and, instead, embraced an accountability model in which organizations must conduct impact assessments and identify and mitigate high risks on an ongoing basis, subject to ex post enforcement.[9]

2.7 The growing role of internal and external AI review boards. Data or AI review boards or advisory panels can be a useful tool when a high-risk AI application is involved. These internal, or sometimes external, standing committees (whose characteristics could be defined by industry or the AI innovation board) can be consulted according to certain risk indicators or escalation criteria to promote a thoughtful dialogue and consideration of the risks and benefits in relation to high-risk AI projects. They may also have the authority to approve a specific use of AI, impose additional specific safeguards, or refuse further development or use of an AI application. CIPL does not suggest requiring AI review boards, but their use should be taken into account as positive evidence of accountability.

2.8 Demonstrating accountability. CIPL has developed an accountability framework to help organisations design, structure, build, implement and demonstrate their data protection management programs based on the key elements of accountability (Leadership and Oversight, Risk Assessment, Policies and Procedures, Transparency, Training and Awareness, Audit and Monitoring and Response and Enforcement – see Appendix 1). CIPL's work confirms that this approach is also scalable to SMEs.[10] In practice, small organisations tend to calibrate accountability measures differently than larger, multinational organisations, and are often able to do so with more agility. It is important in the AI context, and even more so when an AI application presents a high risk, that all organisations, regardless of their size, put in place the necessary processes and controls that the market, business partners, and users expect them to put in place. Business partners in particular will ask for proper assurances before engaging with them.

2.9 Emerging best practices in Accountable AI applications. Many organisations are proactively starting to use accountability frameworks to address the opportunities, risks, challenges and tensions presented by the use of AI, as well as to comply with relevant laws, including data protection and anti-discrimination laws, and to proactively consider social expectations and ensure customer trust. These emerging best practices are starting to take shape in the form of coherent and comprehensive accountable AI frameworks and technical tools, and are likely to catalyse the development and implementation of best practices by all stakeholders in the AI ecosystem, triggering a "race to the top" effect (see Appendix 3 - Mapping Best Practices in AI

[9] See Recital 89 GDPR: "Directive 95/46/EC provided for a general obligation to notify the processing of personal data to the supervisory authorities. While that obligation produces administrative and financial burdens, it did not in all cases contribute to improving the protection of personal data. Such indiscriminate general notification obligations should therefore be abolished, and replaced by effective procedures and mechanisms which focus instead on those types of processing operations which are likely to result in a high risk to the rights and freedoms of natural persons by virtue of their nature, scope, context and purposes."

[10] See What Good and Effective Data Privacy Accountability Looks Like: Mapping Organisations' Practices to the CIPL Accountability Framework

Governance to the CIPL Accountability Framework). These practices will also enable organisations to promote risk-based organisational accountability as a bridge to developing globally harmonised and interoperable guidelines on AI applications and uses.[11]

2.10 **Rewarding and encouraging accountability.** The regulatory framework should also provide appropriate rewards and encouragements to further stimulate and help accelerate AI accountability and organisational best practices. Such "incentives" could include: linking accountability to external certifications; recognising self-regulatory commitments of organisations that publicly define the AI values and principles they implement along with progress against benchmarks;[12] promoting organisational accountability through Digital Innovation Hubs; using demonstrated accountability as a "licence to operate" by allowing accountable and/or certified organisations greater opportunities to use and share data responsibly to facilitate growth in responsible AI uses; allowing broader use of data in AI for socially beneficial projects; using demonstrated AI accountability as a criterion for public procurement projects or B2B due diligence; and recognising demonstrated AI accountability as a mitigating factor or as a liability reduction factor in the enforcement context.

3. Smart and risk-based regulatory oversight

The following elements describe the essential features of an effective oversight framework:

3.1 **Novel and agile regulatory oversight.** Innovative technologies and uses require modern and flexible regulatory oversight. The oversight of AI practices should be based on the current ecosystem of sectoral and national regulators, but with a strong and streamlined, non-bureaucratic collaboration and consistency. Rarely will a situation occur where the risks generated by an AI application are not already under the oversight of an existing regulator. The existing regulators' AI expertise and acumen should be expanded, rather than creating an additional layer of AI-specific agencies.

3.2 **On demand cooperation through AI regulatory hub.** Oversight and enforcement of AI should be performed through an EU structure or AI regulatory hub, composed of AI experts from different regulators, to enable agile cooperation "on demand" and drive consistent application. This could be set up through collaboration schemes via memoranda of understanding to address cases where several regulators may be competent over a specific AI application.

3.3 **Maintain competence of Data Protection Authorities (DPAs) where AI involves the processing of personal data.** The competence of DPAs, or a Lead DPA, where applicable, and the EDPB in cases where an AI application involves the processing of personal data must be maintained. This will require: a) adapting the current GDPR consistency and cooperation procedures to make them

[11] See, for example, the OECD Principles on AI

[12] See, for instance, https://www.telefonica.com/en/web/responsible-business/our-commitments/ai-principles; https://www.ibm.com/blogs/policy/trust-principles/; https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/; https://ai.google/responsibilities/; https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6; https://www.vodafone.com/what-we-do/public-policy/policy-positions/artificial-intelligence-framework

faster and more agile; b) providing additional resources to upskill the current DPA workforce and recruitment of AI and technology experts; and c) creating a permanent AI working group at the EDPB level.

3.4 Risk-based oversight and enforcement. The competent existing regulators should implement risk-based oversight and enforcement, focusing on areas of high-risk AI, which recognize compliance as a dynamic process and journey, allowing bona fide trial and honest error and constant improvement. This approach would create an efficient way to resolve non-compliance issues by enabling the swifter resolution of cases and may also work better in the "grey areas" of compliance, where AI creates tension with the legal norms and where it may take a longer and a more concerted and collaborative effort to find a solution and improve compliance on the ground.

3.5 Enforcement as a last option. Enforcement should be used as a last resort and regulators should prioritise engagement, collaboration, thought-leadership, guidance and other proactive measures to drive better compliance with AI rules. When enforcement is used, regulators should consider the full range of corrective measures taking into account multiple factors, including the nature and gravity of the infringement, the likelihood and severity of harms on individuals, as well as the existence of accountable AI frameworks and practices. Fines should remain a last-resort option for only the most serious, repetitive infringements cases or those that create real and lasting harm for individuals, groups of individuals, organisations, and/or society.[13]

3.6 Relevance of co-regulatory tools. The risk-based approach to AI oversight should be complemented by a consistent EU-level scheme of voluntary codes of conduct, standards, certifications and labeling to help increase trust by demonstrating that an AI application meets certain criteria that have been assessed by an independent body. These must be designed through consultation with stakeholders and updated regularly based on technological developments and new practices.

3.7 Use of innovative regulatory tools. Finally, the EU AI regulatory framework must provide an explicit statutory basis for innovative regulatory oversight tools based on experimentation, such as regulatory sandboxes. Regulatory sandboxes provide organisations with supervised "safe spaces" for building and piloting innovative AI uses in a reiterative manner. They use open and constructive collaboration with regulators to ensure accountable and trustworthy innovation.[14] Because AI applications may impact several fields, regulatory sandboxes would likely require collaboration between several regulators, such as DPAs, as well as relevant sectoral regulators. Each regulator would keep its own competence, but could exchange views and knowledge, align interpretation on risk assessments, trade-offs and mitigation measures or resolve any conflicts of law.

[13] See CIPL's paper Regulating for Results – Strategies and Priorities for Leadership and Engagement

[14] See CIPL Paper Regulatory Sandboxes in Data Protection – Constructive Engagement and Innovative Regulation in Practice. See, also, the ICO, Datatilsynet and CNIL regulatory sandboxes initiatives. Regulatory sandboxes are also increasingly used in other parts of the world (for instance Singapore FinTech Regulatory Sandbox).

Appendix 1 – CIPL Accountability Framework

Organizations must be able to demonstrate accountability – internally and externally

Accountability is not static, but dynamic, reiterative and a constant journey

Accountability requires comprehensive privacy programs that translate legal requirements into risk-based, verifiable and enforceable corporate practices and controls

Company values and business ethics shape accountability

Appendix 2 - CIPL Table on the Application Threshold of Article 22 GDPR

Decisions Producing Legal Effects	• Decisions affecting the legal status of individuals • Decisions affecting accrued legal entitlements of a person • Decisions affecting legal rights of individuals • Decisions affecting public rights — e.g., liberty, citizenship, social security • Decisions affecting an individual's contractual rights • Decisions affecting a person's private rights of ownership
Decisions Producing Similarly Significant Effects *Some of these examples may also fall within the category of legal effects depending on the applicable legal regime and the specific decision in question*	• Decisions affecting an individual's eligibility and access to essential services — e.g., health, education, banking, insurance • Decisions affecting a person's admission to a country, their citizenship, residence or immigration status • Decisions affecting school and university admissions • Decisions based on educational or other test scoring – e.g., university admissions, employment aptitudes • Decision to categorise an individual in a certain tax bracket or apply tax deductions • Decision to promote or pay a bonus to an individual • Decisions affecting an individual's access to energy services and determination of tariffs
Decisions <u>Not</u> Producing Legal or Similarly Significant Effects *These automated decisions do not typically produce such effects. Instances where they might produce such effects are contextual and should be determined on a case-by-case basis.*	• Decisions ensuring network, information and asset security, and preventing cyber-attacks • Decisions to sandbox compromised devices for observation, restrict their access to or block them from a network • Decisions to block access to malicious web addresses and domains and delivery of malicious emails and file attachments • Decisions for fraud detection and prevention (e.g., anti-fraud tools that reject fraudulent transactions on the basis of a high fraud score) • Decisions of automated payment processing services to disconnect a service when customers fail to make timely payments • Decisions based on predictive HR analytics to identify potential job leavers and target them with incentives to stay • Decisions based on predictive analytics to anticipate the likelihood and nature of customer complaints and target appropriate proactive customer service • Normal and commonly accepted forms of targeted advertising • Web and device audience measurement to ensure compliance with advertising agency standards (e.g., requirements not to advertise foods high in fat, sugar and sodium when the audience consists of more than 25% of children)

Appendix 3 - Mapping Best Practices in AI Governance to the CIPL Accountability Framework

This table outlines examples of accountable AI activities undertaken by selected organisations of different sectors, geographies and sizes based on the CIPL Accountability Framework and against each accountability element. The practices are not intended to be mandatory industry standards, but serve as specific examples that are calibrated based on risks, industry context, business model, size and level of maturity of organisations.

ACCOUNTABILITY ELEMENT	RELATED PRACTICES
Leadership and Oversight	• Public commitment and tone from the top to respect ethic, values, specific principles in AI development, deployment and use • Institutionalized AI processes and decision-making with escalation criteria • AI/ Ethics/ Oversight Boards, Committees (internal or external) - to review risky AI use cases and to continuously improve AI practices • Appointing a board member for AI oversight • Appointing a responsible AI lead, AI officer or AI champion • Setting up an internal interdisciplinary AI board or AI committee • Ensuring inclusion and diversity in AI model development and AI product teams
Risk Assessment	• Algorithmic impact assessment or fairness assessment tools to monitor and continuously test algorithms to avoid human bias, unfair discrimination and concept drift throughout the entirety of AI lifecycles • Ethics impact assessment / human rights impact assessment / Data protection impact assessment • Developing standardised risk assessment methodologies, which take into account the benefits and the likelihood and severity of risk factors on individuals and/or society, level of human oversight involved in individually automated decisions with legal effects as well as their explainability according to context and auditability • Trade-offs documentation (e.g., accuracy—data minimization, security—transparency, impact on few—benefit to society) for high-risk processing as part of the risk assessment • Data quality assessment via KPIs • Data evaluation against the purpose—quality, provenance, personal or not, synthetic, in-house or external sources • Framework for data preparation and model assessment – including feature engineering, cross-validation, back-testing, validated KPIs by business • Working in close collaboration between business and data experts (data analysts, data engineers, IT and software engineers) to regularly assess the needs and accuracy results to ensure that the model can be properly used
Policies and Procedures	• Adopting specific AI policies and procedures on how to design, use or sell AI • Policies on the application of privacy and security by design in AI life cycle • Adoption of white, black and gray lists of AI use • Rule setting the level of verification of data input and output • Pilot testing of AI models before release • Use of protected data (e.g., encrypted, pseudonymised, tokenised or synthetic data) in some models • Use of high quality but smaller data sets • Use of federated AI learning models, considering trade-off with data security and user responsibilities • Special considerations for organisations creating and selling AI models, software, applications • Due diligence/self-assessment checklists or tools for business partners using AI

	• Definition of escalation steps with regard to reporting, governance, and risk analysis • Use of baseline model to assess uplift of advanced models (should be used only if needed, model decision/KPI should consider the model complexity to be avoided) • Ideation phase between all stakeholders (data scientists, business, final user, control functions) where needs, outcomes, validations rules, maintenance, need for explainability, budget, are discussed
Transparency	• Different needs for transparency to individuals, regulators, business partners and internally at the different stages of AI lifecycle based on context • Adequate disclosures communicated in simple, easy to understand manner • Take into account that AI must be inclusive and accessible by those with special needs/disabilities • Set up a transparency trail on explainability of decision and broad workings of algorithm to make the AI system auditable • Explain that it is an AI/ML decision, if possibility for confusion (Turing test) • Provide counterfactual information • Understand customers' expectations and deploy based on their readiness to embrace AI • Implement tiered transparency • From black box to glass box—looking at the data as well as algorithm/model • Aspiration of explainability helps understand the black box and builds trust • Define criteria of deployment of AI technologies within the organisation based on usage scenarios and communicate them to the user • Produce model cards (short documents accompanying AI models to describe context in which model should be used, what is the evaluation procedure) • Data hub for transparency on data governance, data accessibility, data lineage, data modification, data quality, definition, etc.
Training and Awareness	• Data scientist training, including how to avoid and address bias • Cross functional training – privacy professionals and engineers • Ethics and fairness training to technology teams • Uses cases where problematic AI deployment has been halted • Role of "translators" in organizations, explaining impact and workings of AI
Monitoring and Verification	• Capability for human in the loop in design, in oversight, in redress • Capability for human understanding of the business and processes using AI • Capability for human audit of input and output • Capability for human review of individual decisions with legal effects • Monitoring the eco-system from data flow in, data process and data flow out • Reliance on different audit techniques • Reliance on counterfactual testing techniques • Pre-definition of AI audit controls • Internal audit team specialised on AI and other emerging technologies • Processes must allow human control or intervention in the AI system where both technically possible and reasonably necessary • Model monitoring (back-testing and feedback loop) and maintenance process
Response and Enforcement	• Processes and procedures to receive and address feedback and complaints • Redress mechanisms to remedy an AI decision • Redress to a human, not to a bot • Feedback channel