Avoiding Gaps: Getting the Scope of Australia's AI Regulatory Framework Right

The Policy Challenge

To effectively protect Australians against AI-related risk, it is critical to ensure that our regulatory framework covers all significant sources of harm. A key challenge to this is how to regulate general purpose AI (GPAI) models, which are highly capable AI systems that can competently perform a range of distinct tasks, including those unintended or unexpected by their developers. GPAI systems are already being deployed in a range of settings, and increasingly form the foundation of many downstream products and services.

Most commonly, GPAI models are generative AI systems, such as Large Language Models (LLMs) like ChatGPT. **They also include newer agentic AI models that can be given a task and execute on its delivery without supervision.** These "AI assistants" can be instructed as if they were humans to complete tasks, and as such, do not have a clearly defined use.

GPAI raises distinct issues for AI regulation. Their **adaptability makes them hard to assess** using conventional risk models. Their **emergent capabilities can lead to unpredictable threats,** even in mundane settings. Their **use as foundational infrastructure means they can have cascading and systemic consequence**s. And their **opaque nature** limits our ability to explain, predict, or correct their behaviour. Australia's approach AI regulation must recognise these issues and the limits on our ability to assess risk in the context of GPAI models.

The Proposed Solution

Australia's AI regulatory framework must explicitly apply to GPAI models to ensure they are subject to appropriate oversight, regardless of how or where they are used. This will ensure a key source of risk is addressed at its source and help Australia's regulations keep pace with this rapidly developing technology.

This requires:

1. **Ensuring Australia's AI Act is not limited to just 'high-risk' AI but also explicitly covers GPAI models**, regardless of the context in which they are used**;**
2. **Appropriately defining GPAI, focusing on a model's capabilities and ability to be used for a variety of tasks,** not just how it is deployed; and
3. **Tailoring obligations between untested GPAI models with potentially dangerous capabilities and those that have been tested and verified as safe** (where this is possible).

This approach is supported by a broad range of stakeholders (see Box 1). **It also aligns with emerging international best practice**, including in the EU, Canada, and United Kingdom.

Box 1. Support for applying the mandatory guardrails to general purpose AI

Accenture: "*Accenture supports the application of the guardrails to all general-purpose AI (GPAI) models. Effective risk mitigation has dependencies on foreseeability, explainability and transparency. With 'black box' GPAI models, risks may be unforeseeable to developers and explainability may not be available to deployers … This lack of explainability can stop users from fully trusting system outputs.*"

ASIC: "*We support the proposal to apply the mandatory guardrails to all general purpose AI (GPAI) models due to the lack of foreseeability in how these models could be used and their capacity to cause harm on a wide-scale and at speed.*"

ARC Centre of Excellence on Automated Decision-Making and Society: "[as] *we do not know all the possible use-cases of a GPAI in advance, we cannot definitively exclude the possibility that it has high risk uses. It therefore makes sense from a regulatory perspective to treat GPAI as a subset of high-risk AI….*"

ACTU: "*... GPAI models should be considered inherently high risk because they can be applied to a wide range of applications, some not possible to foresee, to a potentially very broad audience at very high speeds.*"

CHOICE: "*CHOICE supports DISR's proposal to, at minimum, categorise all GPAI systems as high-risk and subject to mandatory guardrails. However, consideration should be given to whether certain GPAI platforms, designs, deployments, or uses of GPAI may present unacceptable risks.*"

CSIRO: "*Technically, … it is nearly impossible to reliably create a low-risk GPAI system from high-risk GPAI models. … This makes any AI system based on GPAI models inherently high risk, regardless of the use case, risk assessment, or whether it is claimed to be narrow AI via limiting underlying capabilities. … Thus, it is suggested to change "high-risk AI systems and GPAI models" to "high-risk AI models, high-risk AI systems, GPAI models, and GPAI systems.""*

Financial Services Union: "*General purpose AI should be included in any regulation that is being considered by the Government. This includes the Guardrails and the legislative approach that the Government should be taking on AI.*"

Law Council of Australia: "*We agree that the guardrails should apply to all GPAI models in Australia.*"

Victorian Trades Hall Council: "*Legislate stronger transparency and accountability measures for General Purpose Artificial Intelligence (GPAI) systems. … GPAI models are at the frontier of AI capability and this should be reflected in more detailed regulatory obligations. Any adverse impacts of model drift must be reported to regulators and the public registry without undue delay.*"

About Global Shield Australia

Global Shield Australia is an international advocacy organization dedicated to reducing global catastrophic risk. We advocate for credible and effective regulation of artificial intelligence to reduce its potential for harm and thus ensure that its opportunities can be fully realised.

For more information on this proposal or our work, please contact Devon Whittle at devon.whittle@globalshieldpolicy.org or on 0458 980 372.