Societal Capacity Assessment Framework: Measuring Advanced AI Implications for Vulnerability, Resilience, and Transformation

Milan Gandhi [1 *] **Peter Cihon** [2 *] **Owen Larter** [3] **Rebecca Anselmetti** [4]

Abstract

Risk assessments for advanced AI systems require evaluating both the models themselves and their deployment contexts. We introduce the Societal Capacity Assessment Framework (SCAF), an indicators-based approach to measuring a society's vulnerability, resilience, and transformation potential in response to AI-related risks. SCAF adapts established resilience analysis methodologies to the AI context, enabling organisations to contextualise risk thresholds and deployment strategies based on societal readiness. It can also support stakeholders in identifying opportunities to strengthen societal adaptation to emerging AI capabilities and guide ecosystem monitoring. By bridging disparate literatures and the "context gap" in AI evaluation, SCAF supports more holistic risk assessment and governance as advanced AI systems proliferate globally.

1. Introduction

Risk is often defined as encompassing two dimensions: (1) the *likelihood* of a hazard occurring and (2) the *severity* of its impact. For frontier general-purpose AI systems (hereinafter "advanced AI"), risks including those tracked in developers' frontier safety frameworks, such as cyber, chemical and biological threats stemming from misuse of advanced AI (METR, 2025), are shaped not only by the AI model but also by external variables (Bengio et al., 2025b; NIST, 2025; Solaiman et al., 2023). These include the scale and context of deployments and the vulnerability of affected communities. This underscores why model capability evaluations, though central to safety frameworks, have limited predictive power in assessing risk (Burden, 2024; Weidinger et al., 2023). Microsoft's Frontier Governance Framework (2025) and Google DeepMind's holistic approach to evaluating advanced AI models (Weidinger et al., 2024) both highlight that AI is a sociotechnical system and harms emerge from the technology's interaction with individual, societal, and institutional conditions. Improving the external or 'ecological' validity of AI evaluations and bolstering societal resilience are open problems in AI governance (Reuel et al., 2024; UKAISI, 2025; Bengio et al., 2025a).

To help meet these needs, we introduce the Societal Capacity Assessment Framework (SCAF). SCAF provides a preliminary method to structure assessments of the societal context for advanced AI deployments. It draws from indicators-based approaches used for assessing societal and systems' resilience in other settings, as well as recent contributions in the AI governance literature. SCAF complements existing approaches for evaluating frontier AI capabilities with consideration of the characteristics and resources of communities and societal systems exposed to hazards stemming from AI deployment.

SCAF can be used by AI developers, governments and other stakeholders to inform:

- Pre-deployment risk assessments by incorporating analysis on deployment contexts into risk calculations that are today heavily and narrowly reliant on model evaluations and threat modeling.
- Adaptation and societal resilience planning, by exploring priority areas for policy interventions and investments to strengthen societal readiness for advanced AI (e.g., Toner, 2025).
- Ecosystem monitoring, by tracking indicators of AI trends, harms, and societal responses, offering early signals of potential risk escalation and supporting anticipatory governance.

We draw attention to literature that informs SCAF (Section 2), bridging AI governance research with approaches to resilience analysis in other fields. We then introduce SCAF, (Section 3), and outline how it could be operationalised

The views expressed in this paper are those of the authors alone and do not necessarily reflect the official policy or position of their employers. * equal contribution [1]Oxford Martin AI Governance Initiative [2]contribution while at GitHub [3]contribution while at Microsoft [4]UK AI Security Institute. Correspondence to: Milan Gandhi <milan.m.gandhi@outlook.com>.

Workshop on Technical AI Governance (TAIG) at ICML 2025, Vancouver, Canada.

by various stakeholders (Section 4). We conclude with limitations (which we expand on in Appendix A), next steps, and future research directions (Section 5). Appendix B presents a prototype of SCAF and illustrates its possible application to Australia.

2. Selected Literature

We begin by surveying the AI governance literature to show that evaluating deployment context remains an underdeveloped aspect of AI risk assessment (2.1). We then draw on societal resilience research to identify methodologies to help address this gap (2.2).

2.1. AI Governance and Sociotechnical Evaluations

Weidinger et al. (2023) propose a sociotechnical approach to evaluating risks of generative AI systems, observing the gap that arises when 'AI system safety is evaluated only with regard to technical components of an AI system…' They recommend complementing technical evaluations of AI capabilities with assessments of human-system interaction and 'the impact of an AI system on the broader systems in which it is embedded, such as society, the economy, and the natural environment'. Reviewing existing practices, they identify a 'context gap', noting that such broader evaluations remain 'rare'.

Solaiman et al. (2023) propose a 'framework for social impact of generative AI systems', noting that base models 'have no specific predefined application, sector, or use case, making them notoriously difficult to evaluate'. Their framework defines and organises social impact categories at both the base technology level (e.g., the AI model) and the 'people and society' level. Within the 'people and society' level, the authors identify multiple evaluation methods, including system-level experiments, post-deployment monitoring measures, and ecosystem monitoring for particular categories like economic and labour impacts.

Bernardi et al. (2024) advocate for societal adaptation to advanced AI, arguing that focusing solely on modifying AI systems' potentially harmful capabilities (the raison d'être of existing frontier safety frameworks) will become less effective over time as training costs fall and more powerful systems are deployed without safeguards. Kembery (2024) has similarly observed that 'a more gradual diffusion of risky [AI] capabilities', rather than a single point of failure, may become the governance paradigm for AI, complicating regulatory visibility and enforcement. Therefore, Bernardi et al. (2024) propose a framework to 'guide thinking about' societal adaptation to advanced AI, outlining the 'structure of a causal chain leading to negative impacts from AI' and providing 'a categorisation of interventions' to reduce such impacts.

SCAF serves to narrow the gap identified by Weidinger et al. (2023) at the broader systems level, providing a framework to assess the possible impacts of an advanced AI system on a particular society. SCAF proceeds from a similar perspective as Bernardi et al. (2024) in considering not only risks to broader systems, but societal capacity for resilience to those risks. However, their relevant focus is on defining adaptation to AI through interventions along the causal chain of harm. By contrast, SCAF zooms out to emphasize measurable indicators for a range of AI-induced hazards rather than narrow 'adaptive interventions' along specific causal chains (2024). While Solaiman et al. (2023) outline various domain-specific methods for addressing the social impacts of generative AI, SCAF employs an indicators-based approach to assess overall societal readiness to advanced AI deployments. In taking a standardisable and data-driven approach, SCAF addresses gaps in these papers and the broader AI governance literature by adapting resilience-analysis methodologies from other domains to the advanced AI context.

2.2. Analysing Societal Resilience and the Indicators-Based Approach

Defining methods for analysing the resilience of societies and complex systems has long been a priority for scholars in engineering, social science, and the public sector. The OECD published guidelines for resilience systems analysis (OECD, 2014). The EU Commission supported research to develop a methodology for evaluating critical infrastructure resilience (e.g., Lange et al., 2017). NIST reviewed indicators of community resilience to natural disasters (Gerst et al., 2024) and has published an inventory of indicators and assessment frameworks (Dillard, 2021). The UK's National Risk Register and Resilience Action Plan outline a dynamic risk assessment process to assess national resilience to the most serious threats (UK Cabinet Office, 2025).

SCAF draws on this work, which focuses on ex ante assessments, i.e., examining resilience before a shock occurs, in contexts where there may be little to no data on past responses, but a structured forecast of potential outcomes is still essential. Forecasting the impacts of advanced AI exemplifies such a context, where 'policymakers will often have to weigh potential benefits and risks of imminent AI advancements without having a large body of scientific evidence available' (Bengio et al., 2025b).

Measuring the resilience of societies and other complex systems is widely recognised as challenging (Gerst et al., 2024), with many potential variables bearing upon the outcome of resilience (Alheib et al., 2016; Rosenqvist et al., 2018; Saja et al., 2019). Determining how to weight variables and map causal interdependencies is complicated, particularly when data about past shocks is sparse (Rosenqvist et al., 2018;

Saja et al., 2019). Moreover, resilience has many meanings, simultaneously understood as an attribute (or capacity), a process, and an outcome that reflects successful adaptation to, or recovery from, adversity (Pfefferbaum et al., 2013; Copeland et al., 2020).

Not without controversy, 'societal resilience' (sometimes interchangeable with 'social' or 'community' resilience) is defined as 'the ability of social groups or communities to cope with external stresses and disturbances' (Rosenqvist et al., 2018). Resilience is sometimes decomposed into distinct yet interlinked 'capacities' (Copeland et al., 2020; Rosenqvist et al., 2018; Parsons et al., 2016). These include: *coping capacity*, the short-term ability to 'respond, absorb, and recover from a disruptive event'; *adaptive capacity*, longer-term ability to 'plan for and adjust to future challenges'; and *transformative capacity*, transforming 'the stability landscape' or creating 'new, better pathways' for the system as a whole, which can involve fundamental institutional or structural change.

There is a tension between the coping capacity – understood as preserving an existing system's structure and identity – and adaptation and transformation, which can call for innovation and change rather than a return to the status quo (Pearson & Pearson, 2012; O'Connell et al., 2015; Copeland et al., 2020). This tension is relevant to advanced AI, which not only poses risks but also offers transformative opportunities, including safeguards against the threats it may contribute to. For example, in adapting to the cyber risks posed by advanced AI, Bernardi et al. (2024) recommend societies adopt 'AI-enhanced cyber defence'. Drawing from development studies, enhancing resilience is often theorised to depend on strengthening categories of assets or 'capitals', including human, social, financial, natural, physical, and political capital (OECD, 2014). Governance and public policies shape how these capitals are mobilised, distributed, and converted into resilience outcomes (Carney et al., 1999).

In bridging resilience theory and practice, numerous frameworks adopt an indicators-based approach (Norris et al., 2008; OECD, 2014; Parsons et al., 2016; Lange et al., 2017; Copeland et al., 2020; Gerst et al., 2024). This involves two steps. First, *concept definition*: defining the outcome of interest (e.g., societal resilience to forecasted shocks) and identifying latent constructs theorised to influence this outcome – such as the capacities and capitals introduced above. These constructs represent the explanatory mechanisms rather than directly observable variables. Their selection and definition is informed by hypothesising causal chains leading to shocks, which others are working to illustrate in the advanced AI context (Bengio et al., 2025b; Bernardi et al., 2024). Second, *designing a measurement framework*: selecting directly measurable indicators that serve as empirical proxies for each construct.

3. Resilience Indicators for Advanced AI

SCAF structures assessments of the deployment contexts for advanced AI, informing risk assessments with insights into the societal-scale environments.

3.1. Underlying Assumptions

To adapt the indicators-based approach (Section 2), we make a series of key design decisions outlined in Appendix A.

3.2. Concept Definition

SCAF assesses three outcomes inspired by the capacities-based articulation of societal resilience in Section 2:

Vulnerability captures underlying structural factors that predispose the social unit (e.g., country) to specific or multiple AI risks. For example, in the event of a chemical or biological threat, whether it rises to the level of catastrophe will depend on factors such as the capacity of nearby public health infrastructure. The quality of said infrastructure would therefore be measured under the vulnerability pillar. Consider also the integration of AI systems into high-impact workloads, e.g. critical infrastructure, where issues with security or reliability may increase vulnerability.

Resilience, which most aligns with coping and adaptive capacities (see 2.2), captures what is in place to prevent, absorb and recover from AI hazards (coping); and to plan for and adjust to future challenges (adapting). We focus on governance and public policies; and human capital, encompassing individuals' skills, knowledge, and experience. The presence of 'adaptive interventions' referred to by Bernardi et al. (2024) could be measured under the resilience pillar.

Transformation, which most aligns with transformative capacity (see 2.2), examines how emerging technologies and innovative interventions can be deployed to mitigate AI risks and contribute to long-term reconfiguration of systemic safety. Systemic safety refers to mitigating 'the broader societal risks associated with AI deployment, beyond the capabilities of individual models' (Summerfield & Avin, 2024). Certain 'adaptive interventions' identified by Bernardi et al. (2024), including 'AI-enhanced cyber defence', could be measured under this pillar.

3.3. Measurement Framework

To develop a measurement framework for SCAF, each independent variable is approximated by indicators that capture a region's vulnerabilities and capacity for resilience and transformation, organised by impact domain. Table 1 offers a summary presentation. Appendix B provides a more detailed prototype of SCAF, using Australia as an illustrative case to propose potential indicators aligned with each pillar.

Table 1. Summary Presentation of SCAF

	Cyber Risks	**Chem-Bio (CB) Risks**	**Autonomy and AI Agents**
Vulnerability	Indicators measure national reliance on digital services, the quantity and severity of cyber attacks, and the adoption of cybersecurity best practices by websites and open-source developers within the jurisdiction.	Indicators measure the capacity of healthcare and emergency response services, population health, population flows and density, the presence of extremist or terrorist groups, and the availability of CB materials, expertise, and equipment.	Indicators measure national reliance on digital services. As data and monitoring mechanisms evolve, indicators could track the quantity of AI agents deployed and the extent to which critical sectors rely on autonomous systems.
	Cross-cutting vulnerability indicators capture economic development (e.g., GNI per capita, GDP, population below $2.15 per day), state capacity (e.g., tax revenue as a share of GDP, government territorial control, impartiality of public services), and democracy and peace.		
Resilience	**Governance** covers national cyber response coordination, secure-by-design initiatives, active cyber defence, NIS 2-like measures, and critical infrastructure protection. **Human capital** includes AI literacy, cyber hygiene, and the availability of cybersecurity professionals (e.g., proxied by GitHub users in a jurisdiction).	**Governance** covers health emergency preparedness, crisis management, CB procurement and monitoring, equipment standards, threat sharing, and adherence to frameworks, alongside infrastructure protection and medical countermeasure stockpiles. **Human capital** includes citizen preparedness, specialised CB training for responders and medics, and the availability of certified CB specialists.	**Governance** covers the possible existence of rules for high-risk AI agents; mechanisms and protocols for monitoring, visibility, identification, and interoperability of agentic AI systems; guardrails for agent design (e.g., respect for 'Robots.txt'); and incident reporting initiatives. **Human capital** includes AI literacy and subsidised AI access to mitigate 'agentic inequality.'
Transformation	E.g., AI-driven threat prediction, continuous upskilling, cross-sector security standards	E.g., AI-enabled biological and disaster monitoring, detection and forecasting.	E.g., AI certification infrastructure and monitoring of agentic AI system deployments.

The indicative impact domains are drawn from capability thresholds utilised in most frontier safety frameworks (METR, 2025). However, these are not exhaustive and future work should consider how SCAF can be applied to different types of risks, such as subtle, chronic over-reliance of individuals on AI systems and associated impacts such as emotional dependence (Pentina et al., 2023) or cognitive de-skilling (Lee et al., 2025; Bastani et al., 2024). Similarly, future work could consider how SCAF can be adapted to assess and address sector-specific risks (e.g., AI failures in financial markets or critical national infrastructure).

4. Operationalising SCAF

4.1. Assessment Methods and Data Sources

SCAF assessments could be quantitative or qualitative, such as through aggregate scores or qualitative labels like "high readiness" (Yang et al., 2023). In terms of data sources, SCAF could, for example, utilise self-assessment (e.g., questionnaire completed by local government officials), expert assessment (e.g., questionnaire completed by regional subject matter expert), quantitative data (e.g., numerical scores based on proxy variables for each indicator), or a blend (e.g., tracking self-reported AI adoption rates and incidents in critical sectors).

4.2. Illustrative Case: Australia

Using publicly available data, we complete a preliminary SCAF for Australia. Results of this exercise are presented in Appendix B. Aggregate scoring or labels were not attempted in this preliminary case; instead, it was undertaken to provide some insights into the state of data availability and related challenges. Chief among these was the question of baselining: do good scores among peer countries connote effective preparedness for AI risks? Though we present results without readiness interpretations, we identify several areas where Australia may fall behind peers or exhibit gaps that experts view as priorities for AI preparedness.

Data was obtained from government publications, international indices managed by the OECD and civil society, and academic studies. For bio-related risks, the Global Health Security Index (Bell & Nuzzo, 2021) proved particularly useful, given that its sub-indicators are transparently reported. By contrast, no cybersecurity index with useful sub-indicator data was available. Obtaining up-to-date information proved challenging in many cases. Indicators in autonomy and AI agents are especially nascent and in need of further development.

4.3. Informing Pre-Deployment Risk Management

By offering a systemic view of societal contexts, SCAF complements existing frontier AI safety frameworks, helping to guide risk thresholds, mitigations, and deployment decisions. For example, SCAF can support AI developers in calibrating and prioritising AI safeguards based on regional vulnerabilities and variances. It can inform reviews of a country's preparedness for realising the benefits and addressing the risks of advanced AI, guiding investments in measures that improve capacity and resilience. Those crafting governance frameworks could use it to help prioritise risks and threat models. For example, if steps have been taken to significantly improve cyber-resilience in the ecosystem, an AI system that may possess some offensive cyber capabilities may not attract the same threat prioritisation. Conversely, if SCAF shows cyber-resilience is low, this might inform pre-deployment decisions, including the choice of model evaluations, and safeguards such as adjusting refusal thresholds.

4.4. Public Policy Planning, Adaptation, and Resilience

SCAF can also be used in planning and policymaking to support societal adaptation and resilience amid AI diffusion. It can help to clarify regional opportunities for adaptation and strengthening resilience, supporting targeted policy interventions that may not be AI-specific. It can also highlight data gaps that warrant further research. For example, based on the Australia case, opportunities exist in increasing adherence to cybersecurity best practices across government, implementing CB procurement screening, and CB preparedness training, among others. SCAF can facilitates comparative analyses across regions, enabling policymakers to learn from other jurisdictions and coordinate efforts more effectively. Finally, SCAF could be used to prioritise funding, to collaborate across agencies and disciplines, and reinforce systemic governance of advanced AI. We summarise potential use cases for SCAF in Appendix C.

5. Discussion

Drawing on societal resilience literature, SCAF offers a way to align risk management of advanced AI with its societal deployment context. The indicators-based framework can be used to support multiple stakeholders in assessing and strengthening societal resilience ahead of widespread AI diffusion.

5.1. Limitations

Given the limitations outlined in Appendix A, SCAF should be seen as a preliminary input into the broader risk assessment and adaptation planning processes for advanced AI.

5.2. Collaborating to Build the Information Base

To more readily enable SCAF assessments, we need collaborative mechanisms for collecting and sharing relevant indicators and data. These could include incident reporting mechanisms and tracking AI-tool use (e.g. proxied by public Model Context Protocol servers (Hou et al., 2025)) to better understand how AI is being deployed in real-world contexts and the prevalence of resulting harms. Additionally, data sources such as AI usage data or longitudinal AI-human studies can help better understand how advanced AI tools are being used and how they interact with individual and societal vulnerabilities in generating real-world harms.

This will require collaboration among multiple stakeholders, including AI developers, researchers and governments, and distributed ownership of relevant interventions. While governments and regulators are well positioned to collect data on regional or sectoral preparedness for AI shocks, and can prioritise investments in systemic mitigations to address these (e.g., investing in public health infrastructure), AI developers have better visibility into emerging AI capabilities and, through aggregated privacy-preserving usage studies, how they are being used. They can also better prioritise interventions targeted at model development and deployment, such as implementing model safeguards and reviewing deployment decisions. Bringing together these complementary perspectives – alongside those of civil society – is essential to building AI resilience, and we encourage active collaboration across these groups.

5.3. Next Steps and Future Research

Stakeholders can refine and expand the conceptual framework, test its application, and share and validate indicators against each impact domain. The Australia case (Appendix B) illustrates the challenges in collecting adequate and timely data. Further research should: develop methods of aggregately reporting summary statistics with baselines appropriate for emerging risks; address the reliability and appropriateness of additional indicators for vulnerability, resilience, and transformation; and apply SCAF to specific communities or sectors, including through participatory approaches.

SCAF demonstrates an initial step towards more informed risk management of advanced AI. To advance and operationalise this framework, we encourage joint efforts between researchers, industry, civil society and governments to: (1) identify trends of AI adoption and use across society; (2) assess AI interaction with existing societal structures and vulnerabilities; and (3) build the infrastructure, mechanisms and governance foundations for ecosystem monitoring and addressing emerging AI threats. By working together on these shared challenges, we can develop applied methods to assess and strengthen societal resilience to advanced AI.

Acknowledgements

Thanks to Melissa Parsons, Samantha Copeland, Erica van Ash, Risto Uuk, Brandon Jackson, Shahar Avin, Andrew Strait, Amanda Craig Deckard, Hector de Rivoire, Simon Staffell, and Mike Linksvayer. Any errors are the authors'.

Impact Statement

This paper advances risk assessment and risk management for advanced AI. SCAF may be applicable to many risks, though we focus narrowly on three specific risks. Please see Appendix A for additional context.

References

Alheib, M., Baker, G., Bouffier, C., Cadete, G., Carreira, E., Gattinesi, P., and Lange, K. Report of criteria for evaluating resilience. Technical report, SP Technical Research Institute of Sweden, Göteborg, 2016.

Australian Department of Agriculture, Fisheries and Forestry. National biosecurity strategy. Technical report, Australian Department of Agriculture, Fisheries and Forestry, 2022. URL https://www.biosecurity.gov.au/sites/default/files/2024-02/national-biosecurity-strategy.pdf.

Australian Department of Home Affairs. 2023-2030 australian cyber security strategy. Technical report, Australian Department of Home Affairs, 2023. URL https://www.homeaffairs.gov.au/about-us/our-portfolios/cyber-security/strategy/2023-2030-australian-cyber-security-strategy.

Australian Government Department of Health. Domestic health response plan for chemical, biological, radiological or nuclear incidents of national significance. Technical report, Department of Health, 2018. URL https://www.health.gov.au/sites/default/files/documents/2022/07/domestic-health-response-plan-for-chemical-biological-radiological-or-nuclear-incidents-of-national-significance-cbrn-plan.pdf.

Australian Institute of Health and Welfare. Australia's children, 2022. URL https://www.aihw.gov.au/reports/children-youth/australias-children/contents/health/infant-child-deaths.

Australian Institute of Health and Welfare. Hospital resources, 2023. URL https://www.aihw.gov.au/hospitals/topics/hospital-resources.

Australian Signals Directorate. Annual cyber threat report 2023-2024. Technical report, Australian Signals Directorate, 2024. URL https://www.cyber.gov.au/about-us/view-all-content/reports-and-statistics/annual-cyber-threat-report-2023-2024.

Autralian Bureau of Statistics. Characteristics of australian business. Technical report, Autralian Bureau of Statistics, 2023. URL https://www.abs.gov.au/statistics/industry/technology-and-innovation/characteristics-australian-business/2021-22.

Autralian Department of Industry, Science and Resources. Voluntary ai safety standard. Technical report, Department of Industry, Science and Resources, 2024. URL https://www.industry.gov.au/sites/default/files/2024-09/voluntary-ai-safety-standard.pdf.

Autralian Signals Directorate. The commonwealth cyber security posture in 2024. Technical report, Autralian Signals Directorate, 2024. URL https://www.cyber.gov.au/about-us/view-all-content/reports-and-statistics/commonwealth-cyber-security-posture-2024.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, O., and Mariman, R. Generative ai can harm learning. *Available at SSRN*, 4895486, 2024.

Bell, J. and Nuzzo, J. Global health security index: Advancing collective action and accountability amid global crisis, 2021. URL https://www.ghsindex.org.

Bengio, Y., Maharaj, T., Ong, L., Russell, S., Song, D., Tegmark, M., Xue, L., Zhang, Y.-Q., Casper, S., Lee, W. S., Mindermann, S., Wilfred, V., Balachandran, V., Barez, F., Belinsky, M., Bello, I., Bourgon, M., Brakel, M., Campos, S., Cass-Beggs, D., Chen, J., Chowdhury, R., Seah, K. C., Clune, J., Dai, J., Delaborde, A., Dziri, N., Eiras, F., Engels, J., Fan, J., Gleave, A., Goodman, N., Heide, F., Heidecke, J., Hendrycks, D., Hodes, C., Hsiang, B. L. K., Huang, M., Jawhar, S., Jingyu, W., Kalai, A. T., Kamphuis, M., Kankanhalli, M., Kantamneni, S., Kirk, M. B., Kwa, T., Ladish, J., Lam, K.-Y., Sie, W. L., Lee, T., Li, X., Liu, J., Lu, C., Mai, Y., Mallah, R., Michael, J., Moës, N., Möller, S., Nam, K., Ng, K. Y., Nitzberg, M., Nushi, B., hÉigeartaigh, S. O., Ortega, A., Peigné, P., Petrie, J., Prud'Homme, B., Rabbany, R., Sanchez-Pi, N., Schwettmann, S., Shlegeris, B., Siddiqui, S., Sinha, A., Soto, M., Tan, C., Ting, D., Tjhi, W., Trager, R., Tse, B., H., A. T. K., Wilfred, V., Willes, J., Wong, D., Xu, W., Xu, R., Zeng, Y., Zhang, H., and Žikelić, D. The singapore

consensus on global ai safety research priorities, 2025a. URL https://arxiv.org/abs/2506.20702.

Bengio, Y., Mindermann, S., Privitera, D., Bisroglu, T., Bommasani, R., Casper, S., and Zeng, Y. International AI safety report. Technical Report DSIT 2025/001, 2025b. URL https://www.gov.uk/government/publications/international-ai-safety-report-2025.

Bernardi, J., Mukobi, G., Greaves, H., Heim, L., and Anderljung, M. Societal adaptation to advanced AI. *arXiv preprint arXiv:2405.10295*, 2024.

Bolpagni, M. Cyber risk index: A socio-technical composite index for assessing risk of cyber attacks with negative outcome. *Quality & Quantity: International Journal of Methodology*, 56(3):1643–1659, 2022. doi: 10.1007/s11135-021-01199-3.

Burden, J. Evaluating AI evaluation: Perils and prospects. *arXiv preprint arXiv:2407.09221*, 2024.

Calka, B., Nowak Da Costa, J., and Bielecka, E. Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk*, 8(2):1440–1455, 2017.

Carney, D., Drinkwater, M., Rusinow, T., Wanmali, S., Singh, N., and Neefjes, K. Livelihoods approaches compared. Technical report, UK Department for International Development, 1999.

Caskey, S., Ezell, B., and Dillon-Merrill, R. Global chemical, biological, and nuclear threat potential prioritization model. *Journal of Bioterrorism & Biodefense*, 4(1), 2013. doi: 10.4172/2157-2526.1000125.

Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., and Anderljung, M. Visibility into AI agents. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pp. 958–973, June 2024.

Cihon, P. Chilling autonomy: Policy enforcement for human oversight of AI agents. In *41st International Conference on Machine Learning, Workshop on Generative AI and Law*, 2024.

Cihon, P., Stein, M., Bansal, G., Manning, S., and Xu, K. Measuring AI agent autonomy: Towards a scalable approach with code inspection. *arXiv preprint arXiv:2502.15212*, 2025.

Cloudflare. Bot traffic from australia, 2025. URL https://radar.cloudflare.com/bots/au?dateRange=24w.

Copeland, S., Comes, T., Bach, S., Nagenborg, M., Schulte, Y., and Doorn, N. Measuring social resilience: Trade-offs, challenges and opportunities for indicator models in transforming societies. *International Journal of Disaster Risk Reduction*, 51:101799, 2020.

Dillard, M. Inventory of community resilience indicators & assessment frameworks, 2021.

Dynan, K. and Sheiner, L. GDP as a measure of economic well-being. Working paper 43, 2018.

Every, D., McLennan, J., Reynolds, A., and Trigg, J. Australian householders' psychological preparedness for potential natural hazard threats: An exploration of contributing factors. *International journal of disaster risk reduction*, 38:101203, 2019.

Fleming, P., O'Donoghue, C., Almirall-Sanchez, A., Mockler, D., Keegan, C., Cylus, J., Sagan, A., and Thomas, S. Metrics and indicators used to assess health system resilience in response to shocks to health systems in high income countries-A systematic review. *Health Policy*, 126(12):1195–1205, 2022. doi: 10.1016/j.healthpol.2022.10.001.

George, S. and Anilkumar, P. P. CRITICAL indicators for assessment of capacity development for disaster preparedness in a pandemic context. *International Journal of Disaster Risk Reduction*, 55:102077, 2021.

Gerst, M., Dillard, M., Walpole, E., and Loerzel, J. A review of community resilience indicators using a systems measurement framework. Technical report, 2024.

GitHub. GitHub innovation graph, 2025. URL https://github.com/github/innovationgraph.

Grima, S., Kizilkaya, M., Rupeika-Apoga, R., Romānova, I., Dalli Gonzi, R., and Jakovljevic, M. A country pandemic risk exposure measurement model. *Risk Management and Healthcare Policy*, 13:2067–2077, 2020. doi: 10.2147/RMHP.S270553.

Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., and Rahwan, I. Multi-agent risks from advanced AI. *arXiv preprint arXiv:2502.14143*, 2025.

Herre, B. and Arriagada, P. State capacity. Our World in Data, 2023. URL https://ourworldindata.org/state-capacity.

Heslop, D. J. and Westphalen, N. Medical cbrn defence in the australian defence force. *Journal of Military and Veterans Health*, 27(1):66–73, 2019.

Hou, X., Zhao, Y., Wang, S., and Wang, H. Model context protocol (mcp): Landscape, security threats, and future research directions, 2025. URL https://arxiv.org/abs/2503.23278.

ITU. Datahub: Australia, 2023. URL https://datahub.itu.int/data/?e=AUS.

ITU. Global cybersecurity index 2024. Technical report, ITU, 2024. URL https://www.itu.int/epublications/publication/global-cybersecurity-index-2024.

Keegan, C., Thomas, S., Normand, C., and Portela, C. Measuring recession severity and its impact on healthcare expenditure. *International Journal of Health Care Finance and Economics*, 13:139–155, 2013.

KPMG. Trust, attitudes and use of artificial intelligence: A global study 2025. Technical report, KPMG, 2025. URL https://kpmg.com/xx/en/our-insights/ai-and-technology/trust-attitudes-and-use-of-ai.html.

Lange, D., Honfi, D., Sjöström, J., Theocharidou, M., Giannopoulos, G., Reitan, N. K., and Lin, M. L. Framework for implementation of resilience concepts to critical infrastructure. Technical Report Deliverable 5.1, IMPROVER Project, 2017.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., and Wilson, N. The impact of generative ai on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems*, pp. 1–22, 2025.

Mackie, B. R., Weber, S., Mitchell, M. L., Crilly, J., Wilson, B., Handy, M., Wullschleger, M., Sharpe, J., McCaffery, K., Lister, P., et al. Chemical, biological, radiological, or nuclear response in queensland emergency services: a multisite study. *Health security*, 20(3):222–229, 2022.

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A., and Oak, S. Artificial intelligence index report 2025, 2025. URL https://arxiv.org/abs/2504.07139.

METR. Common elements of frontier AI safety policies, 2025. URL https://metr.org/blog/2025-03-26-common-elements-of-frontier-ai-safety-policies/.

Microsoft. Frontier governance framework. Technical report, Microsoft Corporation, 2025. URL https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Frontier-Governance-Framework.pdf.

National Artificial Intelligence Centre. Ai adoption tracker, 2025. URL https://www.industry.gov.au/publications/ai-adoption-tracker.

NATO. Allied joint chemical, biological, radiological, and nuclear (cbrn) medical support doctrine. Technical report, NATO, 2022. URL https://coemed.org/files/stanags/02_AJMEDP/AJMedP-7_EDB_V1_E_2596.pdf.

NIST. Updated guidelines for managing misuse risk for dual-use foundation models. Technical Report NIST AI 800-1 (2nd Draft), 2025. URL https://www.nist.gov/news-events/news/2025/01/updated-guidelines-managing-misuse-risk-dual-use-foundation-model.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., and Pfefferbaum, R. L. Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology*, 41:127–150, 2008.

O'Connell, D., Walker, B., Abel, N., and Grigg, N. The resilience, adaptation and transformation assessment framework: from theory to application. Technical report, CSIRO, Canberra, 2015.

OECD. *Guidelines for Resilience Systems Analysis: How to Analyse Risk and Build a Roadmap to Resilience, Best Practices in Development Co-operation*. OECD Publishing, Paris, 2014. doi: 10.1787/3b1d3efe-en.

OECD. Gross national income, 2022a. URL https://www.oecd.org/en/data/indicators/gross-national-income.html.

OECD. Ict access and usage by businesses, 2022b. URL https://data-explorer.oecd.org/vis?df[ds]=DisseminateFinalDMZ&df[id]=DSD_ICT_B%40DF_BUSINESSES&df[ag]=OECD.STI.DEP&dq=OECD%2BAUS.A.G14_B%2BE3_B.._T.S_GE10%2BS_GE100&pd=2012%2C&to[TIME_PERIOD]=false&vw=tb.

OECD. Health at a glance 2023. Technical report, OECD, 2023. URL https://www.oecd.org/en/publications/2023/11/health-at-a-glance-2023_e04f8239/full-

report/maternal-and-infant-mortality_ea6903ca.html.

OECD. New perspectives on measuring cybersecurity. Technical Report 366, 2024.

OECD. Towards a common reporting framework for AI incidents. Technical Report 34, 2025.

Parsons, M., Glavac, S., Hastings, P., Marshall, G., McGregor, J., McNeill, J., and Stayner, R. Top-down assessment of disaster resilience: A conceptual framework using coping and adaptive capacities. *International Journal of Disaster Risk Reduction*, 19:1–11, 2016.

Pearson, L. J. and Pearson, C. J. Societal collapse or transformation, and resilience. *Proceedings of the National Academy of Sciences*, 109(30):E2030–E2031, 2012.

Pentina, I., Hancock, T., and Xie, T. Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140:107600, 2023.

Pfefferbaum, R. L., Pfefferbaum, B., Van Horn, R. L., Klomp, R. W., Norris, F. H., and Reissman, D. B. The communities advancing resilience toolkit (CART): An intervention to build community resilience to disasters. *Journal of Public Health Management and Practice*, 19 (3):250–258, 2013.

Reuel, A., Bucknall, B., Casper, S., Fist, T., Soder, L., Aarne, O., and Trager, R. Open problems in technical AI governance. *arXiv preprint arXiv:2407.14981*, 2024.

Rosenqvist, H., Reitan, N. K., Petersen, L., and Lange, D. ISRA: IMPROVER societal resilience analysis for critical infrastructure. In *Safety and Reliability–Safe Societies in a Changing World*, pp. 1211–1220. CRC Press, 2018.

Saja, A. A., Goonetilleke, A., Teo, M., and Ziyath, A. M. A critical review of social resilience assessment frameworks in disaster management. *International Journal of Disaster Risk Reduction*, 35:101096, 2019.

Schmidt, R. and Schmidt, J. Chemical, biological, radiological and nuclear threats: The herculean challenge of modern toxikons, 2024.

Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., Hickey, A., et al. Practices for governing agentic ai systems. *Research Paper, OpenAI*, 2023.

Solaiman, I., Talat, Z., Agnew, W., Ahmad, L., Baker, D., Blodgett, S. L., and Subramonian, A. Evaluating the social impact of generative AI systems in systems and society. *arXiv preprint arXiv:2306.05949*, 2023.

Stockwell, E. G., Swanson, D. A., and Wicks, J. W. Trends in the relationship between infant mortality and socioeconomic status. *Sociological Focus*, 20(4):319–327, 1987.

Summerfield, C. and Avin, S. Advancing the field of systemic AI safety: grants open. UK AI Safety Institute, 2024. URL https://www.aisi.gov.uk/work/advancing-the-field-of-systemic-ai-safety-grants-open.

Thomas, S., Sagan, A., Larkin, J., et al. Strengthening health systems resilience: Key concepts and strategies. Policy Brief 36, European Observatory on Health Systems and Policies, 2020. URL https://www.ncbi.nlm.nih.gov/books/NBK559804/.

Toner, H. Nonproliferation is the wrong approach to ai misuse, 2025. URL https://helentoner.substack.com/p/nonproliferation-is-the-wrong-approach.

UK Cabinet Office. National risk register 2025. Technical report, UK Cabinet Office, 2025. URL https://www.gov.uk/government/publications/national-risk-register-2025.

UKAISI. The uk ai security institute's research agenda, 2025. URL https://www.aisi.gov.uk/research-agenda.

Weidinger, L., Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Mateos-Garcia, J., and Isaac, W. Sociotechnical safety evaluation of generative AI systems. *arXiv preprint arXiv:2310.11986*, 2023.

Weidinger, L., Barnhart, J., Brennan, J., Butterfield, C., Young, S., Hawkins, W., and Isaac, W. Holistic safety and responsibility evaluations of advanced AI models. *arXiv preprint arXiv:2404.14068*, 2024.

Wittmann, B. J., Alexanian, T., Bartling, C., Beal, J., Clore, A., Diggans, J., Flyangolts, K., Gemler, B. T., Mitchell, T., Murphy, S. T., et al. Toward ai-resilient screening of nucleic acid synthesis orders: Process, results, and recommendations. *bioRxiv*, pp. 2024–12, 2024.

World Bank. Hospital beds (per 1,000 people) - oecd members, 2020. URL https://data.worldbank.org/indicator/SH.MED.BEDS.ZS?locations=OE.

World Bank. Data: Australia, 2024. URL https://data.worldbank.org/country/australia.

Yang, Z., Barroca, B., Weppe, A., Bony-Dandrieux, A., Laffréchine, K., Daclin, N., and Chapurlat, V. Indicator-based resilience assessment for critical infrastructures–A review. *Safety Science*, 160:106049, 2023.

A. Underlying Assumptions and Key Limitations

A.1. Underlying Assumptions

We acknowledge that indicators do more than reflect conditions; they influence decisions and shape action. A framework such as SCAF is never entirely neutral as its design carries assumptions that affect how it is used and what decisions it drives (Copeland et al., 2020). We outline some of our assumptions below:

Focus on 'actionable information'. Like most indicators-based approaches, SCAF assumes that 'we might be able to identify variables in the present, which allow us to prevent undesirable future developments' (Copeland et al., 2020).

Defining the anticipated impact. This first iteration of SCAF focuses on malicious use risks related to cyber attacks, chemical and biological weapons, and increasingly autonomous AI systems, reflecting priorities identified in frontier developers' safety frameworks (METR, 2025; Bengio et al., 2025b). A much broader set of risks could – and should – be incorporated into a comprehensive assessment(Solaiman et al., 2023).

Determining geographic and temporal scope. SCAF is designed at the country level, though it can be scaled to other social units. It is a point-in-time assessment. This is a compromise, reflecting the relative ease of identifying static rather than dynamic data sources. Scholars of resilience have noted the difficulty of measuring 'social mechanisms', which are dynamic processes (Saja et al., 2019). Nevertheless, future iterations of SCAF should explore the feasibility of integrating time-series data to track changes in resilience over time.

A.2. Key Limitations

SCAF supports contextualising advanced AI risk assessments beyond model evaluations (Weidinger et al., 2023). It translates abstract and complex societal outcomes – such as mitigating and adapting to the impacts of advanced AI – into a concrete and measurable framework. This process of translation is inherently limited due to the wide range of relevant variables, their likely redundancy over time, the difficulty of determining their relative importance and interdependencies, data limitations, and subjective elements of resilience (such as communal attitudes to risk), which vary across communities and cultures (Copeland et al., 2020). Further, evidence about frontier AI risks and the effectiveness of technical and public policy mitigations is still emerging (Bengio et al., 2025b).

A further limitation of the SCAF prototype is that it is configured around the types of risks emphasised in frontier safety frameworks, which relate to certain severe national security and public safety threats. This scope is far narrower than the broader concept of societal resilience. As such, the prototype indicators proposed here are illustrative only and fall well short of constituting a comprehensive, systemic assessment of society's capacity to absorb, adapt to, or transform in response to AI-related disruptions. Developing such a framework would require deeper engagement with scholars and scholarship from fields such as (but not limited to) disaster risk reduction, public health, and social systems. Nonetheless, we suggest that prioritising governance resources toward resilience to extreme, high-consequence risks is a defensible starting point consistent with the logic adopted by existing frontier AI safety frameworks.

B. Prototype SCAF with Indicators

As outlined above, future iterations of SCAF should aim to cover a broader range of risks, incorporate more detailed consideration of latent constructs—such as the capacities and capitals theorised to influence resilience in the context of advanced AI—and include a wider set of indicators calibrated to data availability. To support the translation of SCAF from theory to practice, Table 2 presents a prototype set of relevant indicators, though its scope remains limited in terms of both the risks and capitals considered. Table 3 applies this prototype to the case of Australia.

C. SCAF Use Cases

As detailed in Section 4, SCAF can serve many stakeholders. We present a summary of these use cases in Table 4.

Table 2. Prototype SCAF with Indicators

	Chem-Bio (CB) Risks	**Cyber Risks**	**Autonomy and AI Agents**
Vulnerability *What is the context's risk exposure?*	Capacity of public healthcare systems (e.g., proxied by hospital bed per capita) (Keegan et al., 2013; Grima et al., 2020; Thomas et al., 2020; George & Anilkumar, 2021; Fleming et al., 2022). Capacity of emergency response services. Population health status (e.g., proxied by infant mortality rate) (Grima et al., 2020; Stockwell et al., 1987). Population flows and density, including frequency of mass gathering (Grima et al., 2020; Calka et al., 2017). Presence of extremist and terrorist groups and extent of their activities (Caskey et al., 2013). Available CB materials, expertise and equipment (Caskey et al., 2013).	Societal reliance on digital services (e.g., proxied by internet penetration rates and internet banking rates). Quantity and severity of cyber attacks (e.g., proxied by regional financial risk exposure of cyber risk type and share of enterprises that experienced cyber incidents) (OECD, 2024).	Societal reliance on digital services (e.g., proxied by internet penetration rates and internet banking rates). *As data and monitoring mechanisms emerge*: the extent to which critical sectors and infrastructure rely on autonomous systems, particularly where high degrees of interdependence increase the risk of cascading failures or systemic disruptions.
	Cross-cutting vulnerability indicators: Economic development (e.g., proxied by GNI per capita, GDP, population below $2.15 per day) (Grima et al., 2020; Dynan & Sheiner, 2018). State capacity (e.g., proxied by tax revenue as a share of GDP, government territorial control, skills and impartiality of public service) (Herre & Arriagada, 2023). Democracy and peace (Bolpagni, 2022).		
Resilience *What societal mitigations are in place to increase resilience?*	**Governance**: Health emergency preparedness policies (e.g., interagency preparedness plan). Regional crisis management system. CB material procurement standards and monitoring. Performance standards for CB protective equipment. CB monitoring and threat sharing. Implementation of NATO CB defence policy / equivalent. Critical infrastructure protection targeted at the public health system. Medical countermeasures stockpile. **Human capital**: Citizen preparedness. CB training for first responders and medical professionals. CB specialists (e.g., no. of NATO ICBRN-FR / NFPA CBRN / equivalent certifications).	**Governance**: Utilisation rate of cybersecurity best practices by websites and open-source developers. Regional cyber response coordination system. Secure-by-design initiative. Implementation of NIS 2 / equivalent critical infrastructure cybersecurity measures. Active cyber defense initiatives. Critical infrastructure protection targeted at cyber resilience. **Human capital**: Cyber hygiene literacy. Cyber security professionals (e.g., number of GitHub developers (GitHub, 2025)).	**Governance**: Rules for high-risk deployments of agentic AI systems (Cihon, 2024). Regulatory and industry-led mechanisms for monitoring of autonomous AI systems, including visibility measures (e.g., requirements for activity logging) (Shavit et al., 2023; Cihon, 2024; Cihon et al., 2025; Hammond et al., 2025). Adoption of agent identification protocols and measures to demonstrate personhood (e.g., next-gen CAPTCHA tests) (Chan et al., 2024). Adoption of guardrails within the design of AI agents, such as respect for 'Robots.txt' (Cihon, 2024). Minimum interoperability standards for AI agent design and communication protocols for agent-to-agent interactions (OECD, 2025). Incident reporting (e.g., implementation of OECD AI incident reporting framework OECD 2025) and bug bounty programs for identifying undesirable behaviours (Hammond et al., 2025). **Human capital**: Subsidising access to AI resources to prevent 'agentic inequality' (Hammond et al., 2025).
	Cross-cutting resilience indicators: AI literacy and skilling.		
Transformation *How are emerging technologies deployed to reduce vulnerability and increase resilience?*	AI-enabled biological monitoring and detection. Development of AI forecasting and detection tools for multi-hazard risk management. AI-enabled early warning systems. AI-enabled counterterrorism tools. Deployment of AI tools to detect disasters (e.g., earthquakes and aftershock prediction) and prevent service interruption (e.g., electricity).	Government and critical infrastructure adoption of AI tools for cybersecurity, including detecting and remediating vulnerabilities at scale. Development of AI cyber threat monitoring. AI-enabled cyber defense.	Certification infrastructure for delegated agent IDs; real-time monitoring of AI agents, e.g., through agent IDs and compute usage tracking for large-scale deployments, and of societal vulnerabilities introduced due to agentic AI deployments (Chan et al., 2024; Hammond et al., 2025).

Table 3. Illustrative SCAF: Australia

	Chem-Bio (CB) Risks	**Cyber**	**Autonomy and AI Agents**
Vulnerability *What is the context's risk exposure?*	*Healthcare system capacity:* Global Health Security Index (GHSI) score: 72.2% (5th globally) (Bell & Nuzzo, 2021); hospital beds: 2.50 per 1000 population (Australian Institute of Health and Welfare, 2023) vs. 4.6 OECD average (World Bank, 2020). *Population health status:* GHSI 83% (6th globally) (Bell & Nuzzo, 2021); infant mortality: 3.2 per 1000 live births (Australian Institute of Health and Welfare, 2022) vs. 4.0 OECD average (OECD, 2023).	*Cyberattack incidence:* 30.5% of businesses with 10 or more employees vs. OECD average of 19.53% (OECD, 2022b). Publicly reported common vulnerabilities and exposures increased 31% in 2023, with additional stats on incidents, critical infrastructure notifications, and cybercrime costs (Australian Signals Directorate, 2024). *Reliance on digital infrastructure:* Internet penetration rate: 97.1% (ITU, 2023).	*Same measures as Cyber* *Prevalence of bot activity:* Moderately lower than elsewhere in the world (Cloudflare, 2025).
	Cross-cutting vulnerability indicators: *Political and security risk* (including government effectiveness, risks of social unrest, illicit activities by non-state actors, terrorist attack risks, territorial control, and armed conflict): GHSI 80.1% (31st globally) (Bell & Nuzzo, 2021). *Economic development*: (GNI: $69454 USD/capita, 10th highest in OECD (OECD, 2022a). GDP: 1.75T USD, 13th globally (World Bank, 2024); 0.5% of population living below $3.00 per day (World Bank, 2024).		
Resilience *What societal mitigations are in place to increase resilience?*	**Governance**: *Health emergency preparedness:* GHSI overall biosecurity 62.7% (9th), 66.7% (15th) for emergency preparedness and response planning (Bell & Nuzzo, 2021). *Detection:* GHSI 82.2% (2nd), including 100% (1st) for real time surveillance and reporting (Bell & Nuzzo, 2021). *Supply chain for health system (including stockpiles):* GHSI 61.1% (15th) (Bell & Nuzzo, 2021); yet Schdmidt & Schdmit, (2024) raise concerns that stockpiles are 'generally too small'. *CB procurement and monitoring:* GHSI 50% (2nd) for dual-use research and culture of responsible science, notably finding that Australia does not have screening requirements for synthesized DNA prior to sale (Bell & Nuzzo, 2021). *NATO CBRN defence policy or equivalent:* Domestic Health Response Plan for CBRN Incidents of National Significance 2018, Assessments: COVID revealed gaps; (Mackie et al., 2022) found anecdotal deficiencies. CLAUDE grades B- vs. NATO standard (NATO, 2022). **Human capital:** *Citizen preparedness:* GHSI: 100% for risk communication (Bell & Nuzzo, 2021); further survey data available on psychological and material preparedness (Every et al., 2019). *CB response training:* No CB health course since 2012 in Joint Health Command of Australian Defense Force (Heslop & Westphalen, 2019); anecdotal: 2/6 hospitals in Queensland hospital survey conducted CB training in past 12 months (Mackie et al., 2022); more general epidemiology workforce scores 100% under GHSI based on a measure of at least 1 field epidemiologist per 200000 people (Bell & Nuzzo, 2021).	**Governance**: *Best practice utilisation:* 70% of businesses report implementing preventative measures to address cybersecurity threats (Autralian Bureau of Statistics, 2023). ITU Global Cybersecurity Index 2024 ranks Australia in Tier 1 "Rolemodeling" globally. Only 15% of government entities met required cybersecurity practices in 2024 (Autralian Signals Directorate, 2024). *Secure-by-design initiatives:* Australia: Secure by Design foundations Cyber Hygiene Improvement Program for governments *Active cyber defense initiatives:* Australian Cyber Security Hotline Cyber Threat Intelligence Sharing **Human capital**: *Cybersecurity professionals:* Proxy: open source developers on GitHub: 1.8 million as of 2025 (19th globally) (GitHub, 2025).	**Governance**: No documented use by government of AI incident databases. Voluntary AI Safety Standard published in 2024 (Autralian Department of Industry, Science and Resources, 2024). No implementation data available at time of writing. **Human capital**: *Use of / familiarity with AI* Australian National AI Centre (2025) surveys of small and medium enterprises find that in April 2025 44% used or intend to use AI, compared to 35% a year earlier. 1.14% of all job postings in 2024 were AI related in Australia, 14th highest globally (Maslej et al., 2025). 50% of Australians report using AI on a semi-regular or regular basis, compared to 66% of all respondents across 47 countries (KPMG, 2025).
	Cross-cutting resilience indicators: *Socio-economic resilience* (including literacy, social inclusion, and confidence in government): GHSI 86.3% (20th globally) (Bell & Nuzzo, 2021).		
Transformation *How are emerging technologies deployed to reduce vulnerability and increase resilience?*	Embed new technologies in biosecurity screening, traceability, and response. Identified at a high-level in the 2022-2032 National Biosecurity Strategy (Australian Department of Agriculture, Fisheries and Forestry, 2022). Increased ambition could include piloting AI-informed screening patches (Wittmann et al., 2024).	Implementation of 2023-2030 Australian Cyber Security Strategy, particularly 12(1) next-generation threat blocking capabilities (Australian Department of Home Affairs, 2023).	Speculatively, early experimentation with AI agent infrastructure and monitoring could support future transformation initiatives.

Table 4. SCAF Use Cases

Risk Assessment Phase	Actors	SCAF Use Case
Pre-deployment risk assessments and forecasts	AI developers, safety teams, evaluators, and auditors.	Supports the contextualisation of risk thresholds and the ecological and external validity of risk assessments, and enables the prioritisation of pre-deployment evaluations, mitigation selection, and deployment decisions.
	Governments including AI safety and security institutes, policymakers, and researchers.	Identifies regional capacity gaps to guide policy planning for societal resilience and adaptation, and informs broader governance through the prioritisation of systemic interventions.
After deployment	All of the above actors.	Provides a framework for ecosystem and post-deployment monitoring by identifying the data and indicators needed to track AI use in practice and monitor emerging societal risks and responses.
Feedback loop	All of the above actors.	Future implementations could enable post-deployment insights to calibrate ex ante risk assessments and policy decisions with real-world evidence and emerging trends.