Position: Strategic Resilience Requires the Targeted Maintenance of Human Technical Capacities

Joan Eleanor O'Bryan [1] **David I. Atkinson** [2]

Abstract

As AI systems are deployed across critical infrastructural, decision-making, and economic domains, two trends compound to create a structural risk: humans lose proficiency in tasks delegated to AI (deskilling), and AI develops capabilities which exceed current human expertise (supersession). Existing scholarship on resilience to advanced AI does not address this risk as a strategic problem, potentially resulting in human disempowerment vis-à-vis AI. In this paper, we address the technical requirements behind *strategic resilience*—the ability of a human society to maintain power over critical systems relative to advanced AI—and propose one productive avenue for future research: a Strategic Human Capacity Reserve (SHCR). The SHCR requires further study along three critical dimensions: threat modeling, skill measurement, and policy development.

1. Introduction

As AI systems mature, humans are delegating an increasing share of important technical tasks to AI models—from medical diagnosis (Reverberi et al., 2022) to public service provision (OECD, 2019) to cybersecurity (Li, 2019). This delegation introduces a systemic societal vulnerability: the erosion (Chowdhury et al., 2025) and supersession (Silver et al., 2016) of human expertise, which could result in a near-term future where humans collectively lack the technical know-how to control key elements of society.

AI governance research has not yet adequately conceptualized the structural risk that follows. Work on AI-driven deskilling (e.g., Orrell 2025) generally concerns individual flourishing rather than collective capability, while research on resilience to loss of control correctly focuses on strategic human empowerment (e.g., Somani et al. 2025), but often neglects the problem of technical expertise. To rectify this problem, we address the need for what we term *strategic resilience*—the ability of human societies to maintain or reassert control over critical systems as AI advances—and propose one avenue for future research: a Strategic Human Capacity Reserve (SHCR).

2. Deskilling is a Strategic Concern

Because the future of AI technologies and social responses to them is highly uncertain, "resilience" has emerged as a popular paradigm for AI governance (see Narayanan & Kapoor, 2025; Kembery & Ammann, 2025; Lang et al., 2026). Resilience, understood as "the ability to withstand shocks, adapt, and recover," enables governments to prepare for potential threats from advanced AI, while minimizing those actions which may harm innovation or lock in sub-optimal regulatory regimes (Ramzanali & Sitaraman, 2026; Narayanan & Kapoor, 2025; Winter & Bullock, 2026).

Different threat models call for different approaches to resilience, and interventions that strengthen resilience in one area may weaken it in another. A basic income, for example, might insulate workers from labor-market disruption (Scarcella, 2022), but could also accelerate exit from critical sectors, increasing institutional dependence on AI replacement. Similarly, attempts to harden cybersecurity across firms (Anthropic, 2026) may lead to deeper AI integration, further eroding human skills. Resilience interventions, then, must be designed with multiple goals in mind.

However, current research on resilience to AI deskilling largely neglects the critical goal of *strategic* capacity maintenance. As such, it fails to articulate responses to the structural risk introduced by the erosion and supersession of human technical skills. (For further discussion of the value of the qualifier "strategic" see Appendix B). Most research on AI deskilling is oriented toward individual economic and psychological resilience. Other research, in particular the technological over-reliance literature, focuses primarily on firm- or operator-level deskilling and its mitigation (see Appendix A for a review).

Maintaining human capacity for strategic purposes has

[1]Harvard University, Cambridge, MA, USA [2]Northeastern University, Boston, MA, USA. Correspondence to: Joan O'Bryan <jobryan@fas.harvard.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

precedent. For instance, the U.S. Naval Academy phased out celestial navigation training, then reinstated it in 2015 as insurance against GPS outages and cyberattacks (Prudente, 2015). There are also numerous examples demonstrating the costs of the failure to engage in human capacity-planning, such as the shortage of COBOL-capable engineers during the COVID-19 pandemic (Delaney, 2020) and the expensive and time-consuming reverse-engineering of the FOGBANK material necessary for the W76 warhead (Lillard, 2009).

However, there is only limited research on strategic capacity maintenance in the face of advanced AI. The UK AI Security Institute has recently called for research proposals concerning "critical overreliance," defined as overdependence on AI in key industries resulting in "catastrophic vulnerabilities" (2025). Similarly, in Moon & Boudreaux's (2026) formal model of agency erosion, the authors briefly highlight the value of benchmarking "reversibility capacity," enabling organizations to "assess whether they could restore human decision making" if needed. This would require "maintaining the human expertise, institutional knowledge, and deliberative infrastructure needed to reverse course."

We further this nascent research trajectory by calling for the strategic maintenance and development of human technical capacities, and provide one mechanism for doing so: a Strategic Human Capacity Reserve. Such a proposal would improve societal resilience in the face of advanced AI by ensuring that humans maintain the key capacities needed to control critical systems and infrastructures in the event of a system disruption, such as AI failure or misalignment.

Strategic capacity maintenance complements other AI safety interventions. For instance, stop-gap measures such as kill switches (MacAskill, 2026) are more viable when the social setting supports pausing AI systems—that is, when humans retain the capability to operate without them. Oversight regimes generally presuppose human evaluators capable of judging model output in their domain (Manheim & Homewood, 2025). The SHCR provides the technical capacities that make such interventions tractable.

3. A Research Agenda for the SHCR

A Strategic Human Capacity Reserve is a deliberately maintained portfolio of strategically prioritized human capacities—and the institutions, training pipelines, and practices needed to sustain them—sufficient to operate identified critical systems independently of AI for a defined response window. Capabilities-based planning is a standard tool in defense and emergency preparedness (Keim, 2013); the SHCR adapts it to the question of which human capacities are strategically important to preserve in the AI era and how to do so. In order to develop the SHCR further, we propose the development of the following research agenda, which contains three prongs:

(R1) Threat modeling: which safety measures require which human capacities? Different human capacities are needed to respond to different threats. For instance, responding to AI failure in the medical system may necessitate a different response profile than a loss of control scenario in the cybersecurity domain. Questions within this category include determination of the acceptable degree of AI augmentation of human skills, the appropriate cross-sector and cross-state coordination mechanisms, and determination of ultimate institutional responsibility.

(R2) Measurement: how do we quantify skill reserves? Examples of desired measurements include (a) longitudinal benchmarks of human performance on typically AI-augmented tasks measured under controlled "AI-off" conditions; (b) sector-specific deskilling or supersession indicators (e.g., unaided diagnostic accuracy in medicine, time-to-restart for software teams cut off from coding assistants); and (c) methods for identifying critical skills—those whose loss would be hardest to recover within useful time horizons. This area is well-aligned with prior technical AI governance research, particularly the Societal Capacity Assessment Framework (Gandhi et al., 2025); for instance, the above metrics may be useful vulnerability indicators.

(R3) Policy development: which interventions effectively sustain reserve capacities? Candidates for interventions include firm-level training subsidies, public-sector employment of skill holders, funding for "AI-free" firms, university programs targeted at strategic capacities, private ecosystems of skill development via games and prizes, and "practice days" during which AI systems are taken offline so that humans operate core systems unaided. Each policy response requires further empirical analysis concerning relative efficacy, cost-benefit analyses, and political feasibility, among other areas.

4. SHCR in Practice

In this section, we provide a short case study to illustrate how the SHCR might operate in a particular area. We turn to the domain of cybersecurity, both because of its strategic importance (Suwansukroj, 2026) and because the already evident cybersecurity deskilling (Boamah, 2026) and supersession (Adkins et al., 2025) associated with the use of contemporary AI systems. We emphasize that this is a high-level sketch provided for the purposes of illustration; we suggest that, in practice, this kind of modeling be performed by groups of both AI and subject-matter experts, on a wide range of specific scenarios.

Scenario: In 2032, major cloud providers and critical-infrastructure operators rely on frontier cyber agents to maintain cybersecurity operations. During a geopolitical crisis,

one provider discovers that its agent may have been manipulated through poisoned data, a mistaken model update, or a compromised toolchain, even as anomalous activity spreads across providers and backup systems. Because penetration duration and scope are uncertain, rolling back to the last-trusted model requires sacrificing the majority of active frontline cyberdefenses.

Without the SHCR: Staff suspect something is wrong but have no recent unaided practice and cannot address the matter themselves. They either continue to use a possibly compromised agent or they shut it off and operate at a severely degraded capacity, leaving the firm vulnerable to attacks. Peer firms are also compromised, meaning that the few remaining engineers with the unaugmented capacities to patch defenses cannot help everyone. Moreover, identifying and deploying these engineers is slow and costly, and they must navigate confusing AI-written codebases, which is time-consuming and increases the likelihood of key errors.

Under the SHCR: When compromised agents are suspected, staff turn to firm-level reserves and report the incident to a central body. Firm-level reserves have been maintained via government subsidies and tested on capacity benchmarks, while mandatory emergency training days ensure team coordination, and human readability audits enable system accessibility. Supplemental staff are supplied via the new "National Cyber Reserve." Though the human-only teams cannot match the speed and coverage of the frontier agents they replace, they can establish a reduced trusted operating posture: preserving identity and backup systems, validating emergency patches for the highest-priority services, and determining which parts of the agent stack can be safely reintroduced.

This scenario is only one of many, and explores a limited selection of possible strategies. However, it demonstrates how the SHCR differs from workforce development and orients emergency-planning for the AI-age. The SHCR deliberately preserves unaugmented and complementary skill as a strategic reserve, in the face of opposing economic incentives and with worst-case scenarios in mind.

A research agenda for cybersecurity resilience would further explore the following:

Threat modeling. There has been some relatively general research on the topic of AI-related disaster response and threat modeling (Boudreaux et al., 2025; Wasil et al., 2024; Ortega, 2025; Somani et al., 2025) but comparatively little which directly looks at the role of human capacity in mitigating escalation. Useful examples of threat modeling in the cybersecurity domain include war-gaming case studies (Centre for the Study of Existential Risk, 2024; Smith et al., 2025) and technical reports on frontier capabilities, threats, and vulnerabilities (Rodriguez et al., 2025; AI Security Institute, 2025; Frontier Model Forum, 2026). Further research should consider directly how human capacities alter scenarios, and the implications for policy.

Measurement. Measurement requires standards, assessment, and tracking. A general call for effective resilience measurements has been made by Gandhi et al. (2025). In the cybersecurity domain, standards are largely voluntary and apply at the firm-level (e.g., NIST, 2024); updates happen through slow, periodic revision rather than any standardized procedure keyed to evolving threats. While there are available assessments of human cyber capacities (Carlton et al., 2019) and deskilling (Boamah, 2026), as well as calls to measure human-AI joint capacities (Kulveit et al., 2026), they do not currently inform a measurement of societal vulnerability, much less a human-only reserve policy. Data aggregation and tracking tends to live in the private sphere (Sachpelidis-Brozos et al., 2026; Cattell et al., 2024), although the recent Executive Order mandated a "security clearinghouse" to coordinate vulnerability discovery and remediation efforts (Holland & Knight, 2026). Combining such data with tracking of critical capacities would provide the critical information needed for policy making.

Policy development. There has been research done on relevant potential policies: standards, testing, and regulatory mandates (e.g., Langridge, 2009); institutional and infrastructure development (e.g., Qian et al., 2018; Baker, 2026); financial/market incentives and economic latency (e.g., Cao et al., 2025; Ward-Jackson, 2026); and public awareness campaigns (e.g., Vaňová et al., 2023). Combining research on strategic latency and reserve policies with the cyber-security threat-modeling and measurement agendas described above, supplemented with explicit cost benefit analyses when feasible, would result in a policy agenda for an effective cybersecurity SHRC.

Beyond cybersecurity, similar exercises are required for a number of key areas, including political decision-making, key infrastructure maintenance, and AI-research itself.

5. Conclusion

The literatures on both deskilling and AI resilience both demonstrate a significant lacuna: a failure to consider the key technical human capacities required for effective AI governance. This paper draws attention to this gap, and proposes one viable solution: the Strategic Human Capacity Reserve. Our research agenda calls for further study in three areas: threat modeling, skill measurement, and policy development. While this paper emphasizes the SHCR, other policies for ensuring the strategic maintenance of human capacities are very much needed. The research agenda proposed may generalize to other potential solutions, and we invite others to build on this approach.

Acknowledgments

David Atkinson is supported by a National Science Foundation CISE Graduate Fellowship.

AI Usage Statement

We used large language models for research queries and assistance in drafting this paper. All substantive claims, analysis, and interpretations are the authors' own.

References

Adkins, H., Evron, G., and Schneier, B. Autonomous AI, hacking and the future of cybersecurity. https://www.csoonline.com/article/4069075/autonomous-ai-hacking-and-the-future-of-cybersecurity.html, 2025. CSO Online. Accessed 2026-06-30.

AI Security Institute. Frontier AI trends report. https://www.aisi.gov.uk/frontier-ai-trends-report, 2025. UK AI Security Institute.

Anthropic. Project glasswing, 2026. URL https://www.anthropic.com/glasswing.

Baker, M. Assessing the feasibility and advisability of a civilian cybersecurity reserve. https://www.sei.cmu.edu/blog/assessing-the-feasibility-and-advisability-of-a-civilian-cybersecurity-reserve/, March 2026.

Bernardi, J. et al. Societal adaptation to advanced ai. *arXiv preprint*, 2024. URL https://arxiv.org/pdf/2405.10295.

Boamah, K. Ai-driven threat intelligence adoption in security operations centers: Human skill erosion or analyst augmentation? a comparative case study of resource-constrained socs in emerging markets. *International Journal of Science and Research Archive*, 19:25–033, 05 2026. doi: 10.30574/ijsra.2026.19.3.1226.

Boudreaux, B., Vermeer, M. J. D., Horton, K., and Kalra, N. The case for ai loss of control response planning and an outline to get started. Technical Report PE-A4232-1, RAND Corporation, 2025. URL https://www.rand.org/content/dam/rand/pubs/perspectives/PEA4200/PEA4232-1/RAND_PEA4232-1.pdf.

Cao, Y., Tian, J., and Huang, K. Public–private collaborations in humanitarian relief supplies: Incentive reserve contracts. *Computers & Industrial Engineering*, 203: 111031, 2025. doi: 10.1016/j.cie.2025.111031.

Carlsmith, J. Is power-seeking ai an existential risk? *arXiv preprint*, 2022. URL https://arxiv.org/abs/2206.13353.

Carlton, M., Levy, Y., and Ramim, M. Mitigating cyber attacks through the measurement of non-IT professionals' cybersecurity skills. *Information and Computer Security*, 27(1):101–121, 2019. doi: 10.1108/ICS-11-2016-0088. URL https://www.emerald.com/ics/article/27/1/101/112107/Mitigating-cyber-attacks-through-the-measurement.

Cattell, S., Ghosh, A., and Kaffee, L.-A. Coordinated flaw disclosure for AI: Beyond security vulnerabilities. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, volume 7, pp. 267–280, 2024. doi: 10.1609/aies.v7i1.31635.

Centre for the Study of Existential Risk. Intelligence rising. https://www.cser.ac.uk/work/intelligence-rising/, 2024. AI strategy wargame, University of Cambridge. Accessed 2026-06-30.

Chowdhury, M., Ding, T., and Nwankwo, E. A simple model for ai-induced de-skilling. Poster, 2025. URL https://iclr.cc/virtual/2025/34879. ICLR 2025.

Cummings, M. L. Automation bias in intelligent time critical decision support systems. In *AIAA 1st Intelligent Systems Technical Conference*, pp. 557–562. American Institute of Aeronautics and Astronautics, 2004. doi: 10.2514/6.2004-6313. URL https://maritimesafetyinnovationlab.org/wp-content/uploads/2023/02/Automation-Bias-in-Intelligent-Time-Critical-Deci pdf.

Delaney, J. COBOL Programmers Are Back in Demand. Seriously. https://cacm.acm.org/news/cobol-programmers-are-back-in-demand-seriously/, 2020. Communications of the ACM (News). Accessed 2026-06-30.

Ferdman, A. Practices worth preserving: Knowing when to offload to technology. *Philosophy & Technology*, 38 (3):1–7, 2025. URL https://philpapers.org/rec/FERADI-6.

Frontier Model Forum. Managing advanced cyber risks in frontier AI frameworks. https://www.frontiermodelforum.org/technical-reports/managing-advanced-cyber-risks-in-frontier-ai-fram 2026.

Gandhi, M. M., Cihon, P., Larter, O. C., and Anselmetti, R. Societal capacity assessment framework. *arXiv*

preprint, 2025. URL https://arxiv.org/pdf/2509.22742.

Goddard, K., Roudsari, A., and Wyatt, J. C. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1):121–127, 2012. doi: 10.1136/amiajnl-2011-000089. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3240751/.

Grissinger, M. Understanding human over-reliance on technology. *P&T: A Peer-Reviewed Journal for Formulary Management*, 44(6):320–375, 2019. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6534180/.

Gulyamov, S. S., Umarkhanova, D., Ergashev, A., Musaev, G., Kan, E., and Rodionov, A. The problem of cognitive atrophy in the environment of permanent digital assistance: Developing a methodology for stimulating independent thinking. *IEEE Xplore*, 2025. URL https://ieeexplore.ieee.org/abstract/document/11211979.

Holland & Knight. Executive order on artificial intelligence expands cybersecurity. https://www.hklaw.com/en/insights/publications/2026/06/executive-order-on-artificial-intelligence-expands-cybersecurity, June 2026. Holland & Knight Insights. Accessed 2026-06-30.

Hough, J., Culley, N., Erganian, C., and Alahdab, F. Potential risks of genai on medical education. *BMJ Evidence-Based Medicine*, 30(6):406–408, 2025. doi: 10.1136/bmjebm-2025-114339. URL https://doi.org/10.1136/bmjebm-2025-114339.

Ibrahim, L., Collins, K. M., Kim, S. S. Y., Reuel, A., Lamparth, M., Feng, K., Ahmad, L., Soni, P., El Kattan, A., Stein, M., Swaroop, S., Sucholutsky, I., Strait, A., Liao, Q. V., and Bhatt, U. Measuring and mitigating overreliance is necessary for building human-compatible ai. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2509.08010.

Imagining the Digital Future. Building human resilience for the age of ai. Elon University, 2026. URL https://www.elon.edu/u/news/2026/04/01/building-human-resilience-for-the-age-of-ai/.

Keim, M. E. An innovative approach to capability-based emergency operations planning. *Disaster Health*, 1(1):54–62, 2013. URL https://pubmed.ncbi.nlm.nih.gov/28228987/.

Kembery, E. and Ammann, N. Ai resilience. Website, 2025. URL https://airesilience.net/.

Kulveit, J., Leech, G., Gavenciak, T., and Douglas, R. AI evaluation should work with humans. https://www.gleech.org/files/withhumans.pdf, 2026.

Lang, N., Langione, M., Iyer, S., Das, A., and Martínez, D. Z. For most countries, ai sovereignty is an illusion. resilience is real., 2026. URL https://www.bcg.com/publications/2026/ai-sovereignty-is-an-illusion-resilience-is-real.

Langridge, R. Confronting drought: Water supply planning and the establishment of a strategic groundwater reserve. *University of Denver Water Law Review*, 12(2):295, 2009. URL https://escholarship.org/content/qt15x54613/qt15x54613.pdf.

Li, J. Cyber security meets artificial intelligence: a survey. *Frontiers of Information Technology & Electronic Engineering*, 2019. doi: 10.1631/FITEE.1800573. URL https://doi.org/10.1631/FITEE.1800573.

Lillard, J. FOGBANK: Lost knowledge regained. Nuclear Weapons Journal Issue 2, Los Alamos National Laboratory, 2009. https://cdn.lanl.gov/files/nwj2-09_62c62.pdf.

Liu, J. et al. Overreliance on ai systems and skill degradation risks among operators in critical infrastructure cybersecurity environments, 2025. URL https://www.researchgate.net/publication/400757324_Overreliance_on_AI_Systems_and_Skill_Degradation_Risks_Among_Operators_in_Critical_Infrastructure_Cybersecurity_Environments.

MacAskill, W. What an international project to develop agi should look like. Forethought, January 2026. URL https://www.forethought.org/research/what-an-international-project-to-develop-agi-shou

Manheim, D. and Homewood, A. Limits of safe ai deployment: Differentiating oversight and control. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2507.03525.

Moon, A. and Boudreaux, B. A formal model of how artificial intelligence erodes human agency. Technical Report RR-A4817-1, RAND Corporation, 2026. URL https://www.rand.org/pubs/research_reports/RRA4817-1.html.

Narayanan, A. and Kapoor, S. Ai as normal technology. Knight First Amendment Institute, 2025. URL https://knightcolumbia.org/content/ai-as-normal-technology.

NIST. The NIST Cybersecurity Framework (CSF) 2.0, 2024. National Institute of Standards and Technology CSWP 29. https://doi.org/10.6028/NIST.CSWP.29.

OECD. State of the art in the use of emerging technologies in the public sector. Technical report, OECD, 2019. URL https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/09/state-of-the-art-in-the-use-of-emerging-technologies_2b6dacca/932780bc-en.pdf.

Orrell, B. De-skilling the knowledge economy. Technical report, American Enterprise Institute, 2025. URL http://aei.org/research-products/report/de-skilling-the-knowledge-economy/.

Ortega, A. AI threats to national security can be countered through an incident regime. *arXiv preprint*, 2025. URL https://arxiv.org/pdf/2503.19887.

Oxford Learner's Dictionaries. Definition of strategic, 2020. URL https://www.oxfordlearnersdictionaries.com/definition/american_english/strategic. Online dictionary entry.

Prudente, T. Naval academy reinstates celestial navigation. https://www.militarytimes.com/news/your-military/2015/11/01/naval-academy-reinstates-celestial-navigation/. November 2015. Military Times. Accessed 2026-06-30.

Qian, G., Liu, B., and Wang, Q. Government subsidies, state ownership, regulatory infrastructure, and the import of strategic resources. *Multinational Business Review*, 26(4):319–336, 2018. URL https://www.emerald.com/mbr/article-abstract/26/4/319/281253/Government-subsidies-state-ownership-regulatory.

Ramzanali, A. and Sitaraman, G. Toward a grand strategy for ai resilience. Technical Report PE-A4525-1, RAND Corporation, 2026. URL https://www.rand.org/pubs/perspectives/PEA4525-1.html.

Reuel, A., Soder, L., Bucknall, B., and Undheim, T. A. Position paper: Technical research and talent is needed for effective ai governance. *arXiv preprint*, 2024. URL https://arxiv.org/abs/2406.06987.

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., and Cherubini, A. Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports*, 12(1):14952, 2022. doi: 10.1038/s41598-022-18751-2. URL https://www.nature.com/articles/s41598-022-18751-2.

Rodriguez, M., Popa, R. A., Flynn, F., Liang, L., et al. A framework for evaluating emerging cyberattack capabilities of AI, 2025. URL https://arxiv.org/abs/2503.11917.

Sachpelidis-Brozos, N., Katsaros, E., Radoglou-Grammatikis, P., Kalitsios, G., Sarigiannidis, A., Seritan, G. C., Politis, I., Xenakis, C., Goudos, S., and Sarigiannidis, P. Hacking intelligence: Mapping the anatomy of adversarial threats in artificial intelligence with MITRE ATLAS. *Computer Science Review*, 61: 100923, 2026. doi: 10.1016/j.cosrev.2026.100923. URL https://doi.org/10.1016/j.cosrev.2026.100923.

Scarcella, L. Universal basic income and labor-market disruption. SSRN Working Paper, 2022. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4014798.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. ISSN 1476-4687. doi: 10.1038/nature16961.

Smith, G., Hage, G., Heitzenrater, C., Chessen, M., and Girven, R. S. Infinite potential—insights from the cyber surprise scenario. Research Report RR-A4626-1, RAND Corporation, 2025. https://www.rand.org/pubs/research_reports/RRA4626-1.html.

Somani, E., Friedman, A., Wu, H., Lu, M., Byrd, C., van Soest, H., and Zakaria, S. Strengthening emergency preparedness and response for AI loss of control incidents. Technical Report RR-A3847-1, RAND Corporation, 2025. URL https://www.rand.org/pubs/research_reports/RRA3847-1.html.

Suwansukroj, A. Cybersecurity: A systemic, economic and strategic imperative. https://www.weforum.org/stories/2026/05/cybersecurity-systemic-economic-and-strategic-imp May 2026. World Economic Forum. Accessed 2026-06-30.

Taylor, C., Keller, J., Fanjoy, R. O., and Mendonca, F. C. An exploratory study of automation errors in part 91 operations. *Journal of Aviation/Aerospace Education and Research*, 29(1), 2020. URL https://commons.erau.edu/jaaer/vol29/iss1/2/.

UK AI Security Institute. Navigating the uncharted: Building societal resilience to frontier ai. AISI Blog, 2025. URL https://www.aisi.gov.uk/blog/navigating-the-uncharted-building-societal-resili

Vallor, S. Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future

of character. *Philosophy & Technology*, 28(1):107–124, 2015. doi: 10.1007/s13347-014-0156-9. URL https://link-springer-com.ezp-prod1.hul.harvard.edu/article/10.1007/s13347-014-0156-9.

Vaňová, A., Vitálišová, K., Rojíková, D., and Škvareninová, D. Promotion of biosphere reserves: How to build awareness of their importance for sustainable development? *Environ*, pp. 49–61, 2023. doi: 10.2478/environ-2023-0023.

Ward-Jackson, G. A US economic security latency fund. https://www.chinatalk.media/p/a-us-economic-security-latency-fund, 2026. Accessed 2026-06-30.

Wasil, A., Smith, E., Katzke, C., and Bullock, J. AI emergency preparedness: Examining the federal government's ability to detect and respond to AI-related national security threats. *arXiv preprint*, 2024. URL https://arxiv.org/pdf/2407.17347.

Winter, C. and Bullock, C. Radical optionality: Governing transformative AI under uncertainty. Institute for Law & AI, April 2026. URL https://radical-optionality.ai/. Online essay. Accessed 2026-06-30.

Wu, J., Thorne-Large, J., and Zhang, P. Safety first: The risk of over-reliance on technology in navigation. *Journal of Transportation Safety & Security*, 14(7): 1220–1242, 2022. doi: 10.1080/19439962.2021.1909681. URL https://www-tandfonline-com.ezp-prod1.hul.harvard.edu/doi/abs/10.1080/19439962.2021.1909681.

A. Categorizing Resilience Literature

Reports on human resilience to AI in the workforce emphasize cognitive and emotional adaptation, cross-disciplinary collaboration, pipeline maintenance, and macroeconomic stabilization via instruments such as unconditional basic income (Imagining the Digital Future, 2026; Hough et al., 2025). Scholarship on social and moral deskilling (e.g., Vallor, 2015) and cognitive atrophy (e.g., Gulyamov et al., 2025) seeks to protect "human capacity cultivation" in order to ensure, for example, we don't "risk losing the 'human' in us" (Ferdman, 2025). While such interventions are important, we argue that they leave a crucial gap: they do not consider the *collective* impacts of human skill erosion and supersession.

The technological overreliance literature, predating the rise of generative AI, examines what happens when human operators delegate consequential tasks to automated systems and lose the ability to intervene when those systems fail (e.g., Grissinger, 2019; Wu et al., 2022). Most such work has remained firm- or operator-level: it asks how individual pilots, clinicians, or other technicians degrade under automation (Taylor et al., 2020; Cummings, 2004; Goddard et al., 2012). But some recent work in this tradition begins to gesture beyond the individual operator, referencing harms whose consequences are no longer confined to the particular firm but accrue to the systems those operators defend (see Liu et al., 2025; Ibrahim et al., 2025).

On the other hand, in the literatures which overlap with our conception of strategic resilience, commonly proposed strategies to improve human resilience to advanced AI include funding research into AI risks (Bernardi et al., 2024), developing incident reporting and early warning systems (Bernardi et al., 2024; Narayanan & Kapoor, 2025), and increasing technical expertise in government (Reuel et al., 2024). While the last is directly implicated in our project, it does not yet address concerns about skill erosion and skill supersession.

B. Definitions of Strategic

The qualifier "strategic" is appropriate because this form of resilience is oriented towards the pursuit of one particular end: meaningful human power over critical systems. A society can be highly resilient in the operational sense while also being strategically fragile: AI systems may run all critical infrastructure smoothly, with disruptions absorbed efficiently, while humans lose the ability to reassert control if doing so becomes necessary.

In addition to the aforementioned sense of "planning for a particular end," strategic can be understood in a combative sense, "connected with getting an advantage in a war or other military situation" (Oxford Learner's Dictionaries, 2020). While currently treating AI systems as weakly non-cooperative—as structural pressures (delegation incentives, infrastructure concentration, skill atrophy) actively erode human capacity even absent malice—strategic resilience leaves open the possibility that AIs should be understood in an adversarial sense in the future (Carlsmith, 2022). In this sense, resilience measures are strategic insofar as they entail the development of advantages vis-à-vis an opponent.