

Organizing Against the Algorithm: Collective Response as a Governance Lever for Gradual Disempowerment in South Africa's AI Infrastructure Buildout

Track: Africa — Policy

Abstract

Gradual disempowerment frameworks (Kulveit et al., 2025) describe how incremental AI deployment can erode human influence over economic, cultural, and political systems without any single decisive takeover event — not a robot uprising, but a slow handover of control that no individual decision-maker planned. Most proposed mitigations focus on technical alignment or top-down governance. This paper argues that collective action — labour organizing, civic mobilization, public pressure — is an underexamined bottom-up mechanism through which affected populations can resist disempowerment, and that South Africa's existing organizing infrastructure makes it an unusually well-positioned test case. We analyze Cassava Technologies' Nvidia-backed AI Factory buildout near Johannesburg (\$720m investment, ~3,000 GPUs) as a concrete case where the gap between who bears infrastructure costs (energy, land, displaced labour) and who captures gains (capital, foreign compute partners) creates exactly the conditions under which historical South African collective action has emerged — and identifies why that action has not yet materialized. We propose three policy interventions that convert existing organizing capacity into formal participation in AI governance.

1. Introduction

Gradual disempowerment, as formalized by Kulveit et al. (2025), describes a risk distinct from sudden AI takeover. Societal systems — labour markets, media, governance — incrementally substitute AI for human participation until human influence over those systems erodes, without any individual decision-maker intending that outcome. The mechanism is structural rather than adversarial: competitive incentives and coordination failures drive a gradual handover of control until societal systems decouple from human feedback. Nobody presses a button and takes over; it just happens because everyone is competing and nobody stops to fix the bigger problem.

Most existing responses to this risk are top-down: technical alignment research, or governance frameworks imposed by states or international bodies. Comparatively little work asks a complementary question: what bottom-up mechanisms allow affected populations to resist disempowerment as it happens, rather than waiting for institutions to act on their behalf?

This paper argues that collective action — the organizing, negotiating, and mobilizing through which workers and communities have historically responded to economic disruption — is a candidate mechanism, and that South Africa is a useful test case for two reasons. First, it has dense, currently-active organizing infrastructure (trade union federations, civic movements, a constitutionally protected right to organize). Second, it currently has a concrete, analyzable case of AI infrastructure deployment — Cassava Technologies' AI Factory buildout — where the distributional question (who bears the cost of this infrastructure, who captures its benefits) is live and unresolved.

We are not claiming collective action will definitively resist gradual disempowerment, or that it substitutes for technical and governance solutions. We are claiming it is a real, observable mechanism that current AI safety governance discussion underweights, and that policy can be designed to support it rather than ignore it.

2. Related Work

The gradual disempowerment framework analyzes how incremental AI capability improvements can undermine human influence over large-scale systems — the economy, culture, nation-states — arguing that the alignment of societal systems with human interests has historically been stable only because human participation was necessary for those systems to function. AI deployment is removing that necessity. The original paper's proposed responses concentrate on the supply side: technical research and governance approaches. Subsequent work has extended the institutional analysis but has, to our knowledge, not engaged labour economics or social movement theory as a parallel resistance mechanism.

This is a notable gap. Labour economics and social movement theory have independently studied almost exactly this dynamic — collective response preceding individual adaptation during technological disruption — for over a century, without reference to AI. Our contribution is to bring these two literatures into contact: to ask what gradual-disempowerment theory looks like once you take seriously that affected populations are not passive adapters but historically demonstrated collective actors.

3. Methodology

We use a structured comparative case method. Rather than narrating South African collective action examples illustratively, we evaluate four historical and contemporary cases against a common four-part framework — trigger condition, primary actors, demand structure, and outcome — extracting the conditions under which collective response emerged and what made it more or less effective.

Case	Trigger condition	Primary actor(s)	Demand structure	Outcome
Mineworkers' wage strikes (post-1994)	Concentrated, identifiable employer; visible wage/safety gap	Sectoral trade unions (NUM, AMCU)	Direct, employer-targeted (wages, safety)	Mixed: wage gains in some rounds; violent suppression in others (Marikana, 2012)
Fees Must Fall (2015–2016)	Diffuse cost increase (tuition) with identifiable institutional target	Student bodies, cross-university coalitions	Policy-targeted (subsidy, fee freeze)	Partial: fee freeze conceded; underlying funding model unresolved

Case	Trigger condition	Primary actor(s)	Demand structure	Outcome
G20 Women's Shutdown (2025)	Diffuse harm (safety, economic precarity) tied to visible event (G20 hosting)	Civic coalitions, women's organizations	Attention/agenda-targeted	Achieved visibility; concrete policy change unclear
Cassava AI Factory buildout (2025–present)	Diffuse, automated labour substitution; no single identifiable employer; costs and benefits separated	None yet organized	N/A — not yet articulated	Open

The pattern that emerges across the first three cases: collective response is more likely to form, and more likely to succeed, when there is (a) a visible triggering harm and (b) an identifiable target institution that can plausibly respond to a demand. Response is harder to mobilize when costs are diffuse and the responsible party is unclear — which is exactly the structural problem posed by AI-driven labour substitution, as opposed to a single mine's wage policy or a university's fee schedule.

This is the central methodological contribution of the paper: it is not that collective action has happened before in South Africa, but that the conditions under which it succeeds are identifiable, and AI deployment as currently structured is missing several of them. That is a testable, falsifiable claim. Section 5 proposes specific interventions to manufacture those missing conditions deliberately.

4. Case Study: The Cassava AI Factory, Johannesburg

Cassava Technologies — now backed by equity investment from Nvidia and a partnership with Google — is building a second AI Factory near Johannesburg, alongside its first in Cape Town, with the Johannesburg facility expected to house around 3,000 GPUs in its initial phase as part of a roughly \$720 million investment intended to expand high-performance computing access across the continent. The company frames the project in sovereignty terms: ensuring Africa has its own production of intelligence, with sovereign AI factories that keep intelligence securely within borders, tune models to local languages and cultures, and cultivate local jobs, startups, and economic growth. Cassava has also stated intent to scale the model to Nigeria, Kenya, Egypt, and Morocco, positioning itself as Africa's first NVIDIA Cloud Partner.

This framing is worth taking seriously, not dismissing. Data sovereignty and local capacity-building are genuine goods, and this paper does not argue the project is illegitimate. The point is narrower: the distributional structure of who bears the buildout's costs and who captures its benefits has not been made public, negotiated, or contested by any organized group — and the comparative case analysis in Section 3 suggests this is precisely the condition under which collective response historically fails to form, not because the harm is small, but because it is diffuse and untargeted.

Three concrete distributional questions remain open:

1. Energy. GPU infrastructure at this scale is energy-intensive, in a country with an unresolved grid capacity history. Who absorbs the cost of additional load — the facility, Eskom, or the public through tariffs or load-shedding exposure? This has not been publicly disclosed.
2. Labour. The jobs created by operating an AI Factory (largely technical, datacenter operations) are not obviously the same population whose labour is displaced by the AI capacity the factory produces (administrative, content, analytical work across the economy that becomes automatable). No mechanism currently connects these two groups, and no party is positioned to advocate for the latter.
3. Capital structure. With Nvidia and Google as equity/commercial partners, a meaningful share of long-run value capture sits with foreign infrastructure and model providers, even where compute is physically sovereign. Whether South African workers, civil society, or government have a seat in evaluating that structure is not publicly documented.

None of this means the project is harmful on net. It means the project currently exhibits the exact structural feature — diffuse cost, unclear responsible party, no organized constituency — that Section 3's comparative analysis identifies as the condition under which South African collective response historically fails to materialize, even when underlying harms are real. This is the gap policy can close.

5. Policy Recommendations

Based on the comparative analysis above, we propose three interventions, each targeting one of the missing conditions identified in Sections 3 and 4.

5.1 Mandatory Distributional Disclosure for Large Compute Infrastructure Projects

Modeled on existing environmental impact disclosure requirements, large AI infrastructure investments (above a GPU/capital threshold) operating in South Africa should be required to publish a labour and energy impact assessment — not just job creation projections, but an analysis of which sectors are exposed to displacement risk from the capacity being built. This directly addresses the diffuse, unclear-target problem: it creates a public document that an organizing effort can contest, analogous to how Fees Must Fall had a specific subsidy policy to argue against.

5.2 A Sectoral Early-Warning Mechanism, Jointly Resourced by Government and Labour Federations

Rather than waiting for displacement to be felt before organizing begins — the historical pattern across Section 3's cases — South Africa's existing union federations are well-positioned to monitor AI deployment announcements and model displacement risk before it materializes. This converts a structural disadvantage of AI-driven disruption (it's diffuse and hard to see coming) into a tractable one, using existing institutional capacity rather than building new capacity from scratch. Critically, data governance for this mechanism must remain worker-controlled, not state- or employer-controlled, to prevent the same infrastructure from being used to pre-empt organizing.

5.3 A Formal Civil-Society Seat in Infrastructure Investment Review

Where state-linked entities (e.g., CSIR, as referenced in Cassava's own public statements) are project partners or beneficiaries, government should require a labour/civil-society observer role in the relevant approval or partnership process — not veto power, but standing and information access sufficient to let organizing happen before a project is operational rather than only in response to harms after the fact.

Each proposal directly supplies one of the three missing conditions — visible target, identifiable responsible party, advance notice — that Section 3's comparative case analysis shows are necessary for the collective response pathway to function at all.

6. Limitations and Dual-Use Considerations

This remains conceptual policy analysis, not empirical social science. No interviews were conducted with affected workers, Cassava employees, or policymakers; the case study relies on public corporate and press statements, which likely understate distributional concerns the company has not chosen to disclose. The four-case comparison in Section 3 is illustrative of a pattern, not a statistically powered test — four cases cannot establish causation, only generate a falsifiable hypothesis for further work.

A framework that strengthens collective organizing capacity around AI deployment also carries two misuse risks worth naming directly. First, the early-warning mechanism proposed in 5.2, if poorly governed, could be repurposed to monitor and suppress organizing rather than support it. Any implementation would need worker-controlled data governance. Second, the broader collective-response framing is rhetorically adjacent to populist anti-technology mobilization that may not be evidence-based or proportionate — the same organizing energy this paper treats as a healthy governance mechanism could, in a different configuration, be directed at halting AI capacity that has genuine developmental benefit, as Cassava's sovereignty framing plausibly does. We flag this as a condition any implementation of Section 5 needs to design around.

7. Conclusion and Future Work

Gradual disempowerment is typically framed as something that happens to populations. This paper has argued that South Africa's organizing history, and a live, analyzable case in the Cassava AI Factory buildout, suggest collective action is a real and underexamined mechanism through which populations can resist that process — but only under conditions current AI infrastructure deployment does not provide. Future work should pursue primary research directly with affected communities and unions near the Johannesburg and Cape Town facilities, legal analysis of whether South Africa's labour framework can accommodate AI-driven (as opposed to single-employer) disruption, and extension of the four-condition comparative framework to AI infrastructure cases in other Global South contexts.

References

- Kulveit, J., Douglas, R., Ammann, N., Turan, D., Krueger, D., & Duvenaud, D. (2025). Gradual Disempowerment: Systemic Existential Risks from Incremental AI Development. [arXiv:2501.16946](https://arxiv.org/abs/2501.16946).
- ITWeb. (2026, May). Cassava plans AI factory in Joburg, as Cape Town goes live.
- Connecting Africa. (2026, March). Cassava deploys 'AI Factory' in SA, more countries in pipeline.
- TechAfrica News. (2026, March). Cassava Technologies Launches AI Factory in South Africa, Plans Expansion Across Africa.