Wellbeing Futures: Innovation, Policy and Research

wfipr.com

Research Article

Toward What? Making Wellbeing Operational in the Economics of Transformative AI: A Response to the NBER Research Agenda

John Ricketts[1,*] and Chris D. Beaumont[2]

[1]*AI+Wellbeing Institute, International College of Liberal Arts (iCLA), Yamanashi Gakuin University, Kōfu, Japan; E-mail: ricketts.john@c2c.ac.jp*

[2]*University of Tokyo, Tokyo, Japan*

(*Corresponding Author)

Cite This: https://doi.org/10.65638/2978-882X.2026.2.08

ABSTRACT: Wellbeing is acknowledged but unmeasured in the economics of transformative AI. The NBER research agenda assembles the profession's current thinking on what happens when machines can do most economically valuable work. Its chapters provide models of growth, distribution, mortality risk, market structure and institutional change, and several write outcomes beyond consumption into those models: an open research objective valued in life-years, survival traded against consumption, welfare conditioned on rights and institutions, formal transition paths for wages. What the agenda does not provide is repeated, distribution-sensitive measurement of multidimensional wellbeing through the AI transition, or rules connecting such measurement to the deployment choices institutions are making now. This paper proposes that operational complement. The agenda's own results remove, one by one, the familiar reasons for expecting consumption to move with the rest of a life, which makes the informativeness of consumption a proposition to test rather than assume. The paper takes its unit of observation from Coyle and Poquiz's factory count: transformation is visible in the composition of economic life before it is visible in the aggregates. It reads three of the agenda's models as compositional shifts toward extraction, proposes measuring wellbeing along the transition paths the agenda models, and builds on the published Anti-Fragile Well-Being framework, adding what that framework does not contain: AI-specific institutional indicators with operational definitions, rules for attributing outcomes when AI agents act for people, a schedule of repeated wellbeing observation and explicit appraisal rules, with a worked research design comparing two AI deployment arrangements and its implications for policy and practice. The contribution is a proposed operational research design, not a completed measurement instrument. The argument is applied to economies that import their AI services.

Keywords: Transformative AI, Wellbeing measurement, Economic composition, Extraction, Policy appraisal, Anti-Fragile Well-Being, AI-importing economies, Interface governance

1. INTRODUCTION: ACKNOWLEDGED BUT UNMEASURED

In 2025 and 2026 the National Bureau of Economic Research convened a research agenda on the economics of transformative AI, edited by Ajay Agrawal, Erik Brynjolfsson and Anton Korinek and now published as The Economics of Transformative AI (Agrawal, Brynjolfsson & Korinek 2026). This paper reads thirteen of its sixteen chapters, together with two close companions (Mullainathan 2025, a published lecture; Perrier 2026, a preprint), as a body of work and asks what it leaves undone. A note on the standing of these sources: the chapters belong to an edited conference volume, each published with a discussant's comment, and most also circulate as NBER working papers; they

Received: October 22, 2026
Accepted: November 19, 2026
Published: November 23, 2026

are not journal-refereed articles. Their results are treated here as modelled and conditional, not as established empirical findings, and Section 9.6 sets out which of this paper's own claims rest on existing evidence, which are interpretation and which remain to be tested.

The chapters work from related but not identical definitions. Korinek and Lockwood and Restrepo take transformative AI (TAI) to mean machines able to perform essentially all economically valuable work; Benjamin Jones works instead from a tenfold increase in the rate of progress on a chosen outcome; others focus on task automation or on AGI. This paper uses "transformative AI" as an umbrella for that family and preserves the differences where they matter. From their premises the chapters reason, with care and real technical range, to a set of results. Output becomes linear in compute. Wages converge to the cost of the compute that would replace a worker. The labour share of income converges to zero. Labour and consumption tax bases erode. Markets fill with agents whose objectives their makers cannot yet specify. Truthful information is undersupplied without adequate regulation. Routine knowledge work is displaced, under stated conditions, and human effort retreats toward the frontier of what is known.

It would be wrong to say the agenda is silent about what any of this is for. Coyle and Poquiz discuss welfare rising while measured output falls, time saved, leisure, care work and the need for outcome-oriented measures. Benjamin Jones leaves the objective of R&D open in his formal model, as an outcome that can stand for productivity, health or longevity, and shows how valuing health gains in quality-adjusted life years yields a return on research. Charles Jones writes a model in which consumption is traded against survival, calibrates it with a value of life, and derives a level of mitigation spending: an appraisal framework, not a gesture. Mullainathan discusses appraising algorithmic interventions by the marginal value of public funds. Restrepo, having derived an economy in which much human work is supplementary, gives formal transition paths for wages and then turns to meaningful work, purpose and fulfilment, and to the unevenness and perceived unfairness of those paths. Korinek and Lockwood set out transitional policy instruments. Athey and Scott Morton name education, training and care as a research agenda for absorbing displaced workers. Perrier writes welfare as a function of allocations, autonomy-related rights and institutional conditions, in the capabilities tradition.

Three things are distinct here, and the distinction organises this paper. An outcome can be *written into a model*, as Benjamin Jones's open outcome, Charles Jones's survival term and Perrier's rights are. It can be *measured empirically*, with an instrument, a sampling frame and a frequency of observation. And it can be given a *specified role in policy appraisal*: as an objective, a constraint, a distributional requirement or an explicit trade-off. The agenda acknowledges non-consumption outcomes throughout, writes several into models, and, where it appraises, does so in the currency of mortality risk, life-years or public funds. What it does not do, in any chapter read here, is observe multidimensional wellbeing repeatedly through the transition it models, or connect such observation to the institutional choices about AI deployment that are being made now. The one chapter that works mainly from data, Manning and Aguirre, measures preparedness for possible displacement, a condition for an outcome and not the outcome itself.

> *An outcome can govern a decision without being measured: a right or a precautionary floor does exactly that. What it cannot do without measurement is be assessed once the decision is taken, for its consequences or for compliance. The agenda's acknowledged outcomes are, for the most part, in neither position: not measured, and, outside the mortality and life-year cases, not given a stated role. This paper proposes the measurement design and the roles; it does not report measurements.*

The contribution, stated once so that it can be held to: the agenda provides models of growth, distribution, mortality risk and institutional change; this paper proposes an operational complement, a research design for repeated, distribution-sensitive measurement of multidimensional wellbeing, linked to AI deployment arrangements and to explicit rules for policy appraisal. It is a design, not a completed instrument: no indicator in Section 9 has yet been computed and no observation wave has been run.

What is new relative to the published Anti-Fragile Well-Being framework (Yadava *et al*. 2025) should be stated as plainly. That framework supplies the six wellbeing domains, the narrative layer and the adaptive policy cycle; none of those is claimed here. This paper adds four things the framework does not contain: (i) a reading of the NBER agenda that identifies which conditions for consumption to serve as a proxy the agenda's own models remove, stated as a testable proposition (Section 3); (ii) compositional indicators of AI-related institutional change, each tied to a mechanism in the agenda, given a provisional operational definition and paired with the wellbeing domain it is expected to move (Sections 5 and 9.1, Table **2**); (iii) rules for attributing outcomes when an

AI agent acts for a person (Section 9.2); and (iv) a pre-specified, non-compensatory appraisal rule applied to a comparison of two AI deployment arrangements (Section 9.3). Other work by the authors cited below is either unpublished or recently published. It is used for context, is labelled as such where it appears, and none of the four contributions depends on it.

What follows is the structure. Section 2 summarises the agenda. Section 3 makes the case that such measurement must be central rather than supplementary, and states the proposition on which that case rests. Section 4 takes the unit of observation from Coyle and Poquiz's steam-engine analogy: transformation is visible in composition before it is visible in the aggregates. Sections 5 to 7 read the agenda's models in that unit: three mechanisms by which composition can shift toward extraction, transition paths that are modelled but not yet observed in wellbeing terms, and an application to economies that import their AI services. Section 8 draws the implication for education. Section 9 is where the contribution has to be earned: it builds on a published framework to propose indicators with operational definitions (Table **2**), an observation schedule and appraisal rules, gives a worked research design, separates what is established from what remains to be tested (Section 9.6, Table **3**), and draws the implications for policy and practice (Section 9.7). Section 10 concludes.

This is a completion, not a rejection. The agenda's models are used as found. Where we disagree with their authors, the disagreement is about what to count, and about whether the things they themselves say matter can be left outside the count.

2. WHAT THE AGENDA SAYS

The chapters group naturally. Table **1** gives the core claim of each as we read it.

Two things about the set as a whole. It is mainly theoretical; Manning and Aguirre work from data, and Brynjolfsson and Hitzig offer descriptive evidence. And its data and policy instruments are mostly American, with exceptions that matter: Coyle and Poquiz draw on UK statistics, and Athey and Scott Morton explicitly on importing countries. Section 7 builds on those exceptions.

3. WHY MEASUREMENT MUST BE CENTRAL

The claim of this paper is not that the agenda lacks objectives or appraisal. It is that the agenda's own results remove the familiar reasons for expecting consumption to track the rest of a life, at the same moment the agenda leaves the rest of a life unmeasured through the transition it models. Four results do the work.

Restrepo. Wages converge to the opportunity cost of the compute required to reproduce human work, and the labour share converges to zero. Income decouples from work. Restrepo discusses universal income and treating compute as a public resource whose returns are broadly distributed, and Korinek and Lockwood model human welfare from consumption funded by taxing AI capital. We do not dispute that transfer-financed consumption yields welfare, and transfers can improve security and agency. The point is narrower: when consumption is financed by transfer, it may no longer carry the information about contribution, standing and purpose that earned income carried alongside it. Those were never consumption. Consumption stood in for them because the two moved together, and the model has nothing that keeps them together.

Coyle and Poquiz. The great majority of generative AI use is consumed with no household transaction to record; paying subscribers are a small fraction of users, and user counts in any case say nothing about the volume of service consumed without payment. Coyle and Poquiz propose valuation methods and outcome-oriented complements. We take the proposal seriously: the thing being consumed has to be observed on the outcome side, because the expenditure side cannot see it.

Perrier. When an agent chooses for a principal, the chooser and the welfare subject come apart. Perrier's autonomy-qualified theorem recovers efficiency when attribution, internalisation and verification conditions hold; outside them the efficiency guarantee no longer follows, which is not the same as a demonstration of inefficiency. Whether the conditions hold is an empirical matter. It cannot be settled by revealed preference, because the preference revealed is the agent's.

Korinek and Lockwood. The fiscal shadow of the same facts. A tax base erodes when the thing it was attached to stops tracking the thing that mattered.

These are reasons, not results. Most of the agenda's models do not represent purpose, agency or relationships at all, so they cannot show those dimensions moving apart from consumption. What they establish is that each of the conditions under which consumption has been a serviceable proxy, income earned by work, goods that carry a price, and choices made by the person whose welfare is at stake, is

Table 1: The agenda's chapters by theme, with the core claim of each as read in this paper

Theme	Chapter	Core claim
Capabilities and growth	B. Jones (2026)	AI's effect on R&D depends on the share of research tasks it can do, its productivity at them, and how hard the remaining bottlenecks bind. The outcome of R&D is an open variable in the model, with health gains valued in life-years as a worked case; transformative AI is defined as a tenfold increase in the rate of progress on the chosen outcome.
	Agrawal, Gans and Goldfarb (2026)	Routine workers apply known knowledge; geniuses create new knowledge at a cost rising with distance from what is known. Under stated cost assumptions, AI geniuses push human geniuses toward novel questions in the short run; routine workers are displaced in the long run only if AI becomes efficient enough. The conclusion calls for models of complementary human-AI workflows.
	Restrepo (2026)	Under AGI all bottleneck work is automated, while some supplementary work may remain with people. Output becomes additive in compute and labour, growth is pinned to compute, wages converge to the compute cost of reproducing human work, and the labour share converges to zero. Propositions 7 and 8 give formal transition paths: gradual adjustment when compute binds, possible wage discontinuities when algorithms bind. A closing discussion addresses redistribution, compute as a public resource, and meaning.
Agents and markets	Hadfield and Koh (2026)	AI agents are optimisers, but their objectives are currently opaque; the analogy is the incomplete contract between designer and agent. Well-functioning agent markets will need institutions that do not yet exist.
	Shahidi, Rusak, Manning, Fradkin and Horton (2026)	Agents as market participants cut transaction costs and introduce new frictions; they widen the market-design frontier and raise problems of preference fidelity, privacy, autonomy and consumer protection.
	Perrier (2026)	The First Welfare Theorem assumes a binary between autonomous welfare subjects and passive tools. Post-AGI, autonomy is a spectrum. Welfare is written as a function of allocations, autonomy-related rights and institutional conditions; an autonomy-qualified theorem recovers efficiency under explicit attribution, internalisation and verification conditions, and outside them the efficiency guarantee no longer follows.
Firms and organisation	Chatterji, Rock and Talamàs (2026a), with Ide and Talamàs (2025)	AI breaks the codifiability constraint. Autonomous AI raises output more but favours the most knowledgeable; non-autonomous AI benefits the less knowledgeable. Machine-tacit knowledge appears: what machines know but cannot tell. Open questions follow for firm structure, investment and competition.
	Brynjolfsson and Hitzig (2026)	Hayek's two reasons for decentralisation, inalienable local knowledge and bounded processing, are both weakened by AI. In a property-rights model the balance tips toward centralisation, subject to countervailing forces: limits to codification, reasons for decentralisation that persist even among AI systems, and legal requirements for human control. Descriptive evidence of rising concentration and political concern about reduced investment in people.
Distribution and welfare	Athey and Scott Morton (2026)	Treating AI as a priced, imported factor, market power in the AI stack can lower real wages and aggregate welfare despite efficiency gains; rents leak abroad; a monopolist prices against firms' adoption constraint. The chapter explicitly addresses AI-importing countries and competition authorities beyond the United States.
	Manning and Aguirre (2026)	Exposure is not displacement. Across 356 US occupations covering 95.9% of the workforce, exposure and adaptive capacity are positively correlated (r = 0.50), but 6.1 million workers (4.2% of the sample workforce) are in highly exposed occupations with low adaptive capacity, concentrated in clerical and administrative roles. The index measures preparedness for possible displacement.
	Korinek and Lockwood (2026)	TAI erodes the labour and consumption tax bases. Consumption taxation becomes primary, then inadequate; taxing autonomous AGI is an optimal harvesting problem. Transitional instruments and proactive institutional adaptation are set out.
Information and decisions	Stiglitz and Ventura-Bolet (2026)	Information is a public good. AI and platforms raise processing efficiency but endanger the producer business model, cheapen misinformation and change consumers' ability to screen. Absent adequate regulation, truthful information can decline and polarisation rise.
	Mullainathan (2025)	Economics is about decisions, and algorithms change how decisions are made. They discover information people miss; the author is explicit that this is not an argument for automation, and discusses appraisal by the marginal value of public funds.
Measurement	Coyle and Poquiz (2026)	Official statistics will see AI's transformation only with long lags; free use is missing from household consumption; input-output tables show consequences, not the transformation itself. Time use and outcome-oriented measures are proposed as complements.
Risk	C. Jones (2026)	Trading present consumption against survival under calibrated risk and value-of-life assumptions, spending at least 1% of GDP a year to reduce catastrophic AI risk can be justified under most scenarios considered, even without valuing future generations.

removed in the model. Whether co-movement survives their removal is an open question, and we state it as one.

Proposition. As the share of household resources derived from labour income falls, as exposure to services consumed at a zero price rises, and as the share of economically significant choices delegated to AI agents rises, the information that measured consumption carries about the non-material dimensions of wellbeing (agency, purpose, security, relationships, time use) may decline.

"Information" is defined operationally as out-of-sample predictive performance: for each wellbeing domain, a consumption-only prediction compared with a richer model, across bands of labour-income share, free-service exposure and delegation. A decline in predictive association is a finding; that AI caused it is a separate claim requiring its own design. Section 9.5 specifies the panel study.

Material consumption remains one dimension of a life, and a real one. The others are the ones the proposition says may come loose from it. A measure that observes only the first is not wrong about the first; it is blind to the movement the agenda gives reasons to look for.

Three further considerations make measurement central rather than supplementary.

The appraisal argument. An outcome can enter policy appraisal in four ways: as an objective to be maximised, as a constraint that must be respected, as a distributional requirement about who may bear what, or as an explicit trade-off against other outcomes. Rights and precautionary constraints can govern decisions under incomplete measurement, and the agenda shows appraisal done well in its own currency: Charles Jones's survival trade-off, Benjamin Jones's return on life-years, Mullainathan's marginal value of public funds. What measurement adds is the ability to assess consequences and compliance after the fact, and to let a dimension that is currently only acknowledged take one of the four roles with evidence behind it. For the dimensions of a lived transition, agency, purpose, security, relationships, time, the agenda specifies neither the role nor the evidence. Until both are specified, a deployment decision that trades measured cost against unmeasured meaning is settled before it is made, not because anyone is indifferent to meaning but because nothing has been given to weigh. A dashboard does not settle the question of role; it makes the question answerable.

The timing argument. Coyle and Poquiz's central warning is that the aggregates lag. If the transformation is visible in output only with long delay, a measurement programme built on output reports the transition after it has happened. Direct observation of wellbeing at monthly to annual frequency is observation available while the transition is under way, and Restrepo's transition paths give it something to be observed alongside. Manning and Aguirre show what observing during the transition looks like for one conditioning variable; no chapter yet does it for an outcome.

The feasibility argument. Wellbeing measurement is not speculative. National statistical offices in the UK, New Zealand and elsewhere collect it; the OECD (2025) has just updated its guidelines with survey modules for life evaluation, affect and eudaimonia, with implementation guidance; the World Happiness Report compares it; and the literature knows its problems, including adaptation, differences in scale use across cultures and the gap between evaluative and experiential measures. These are real and known. The proxy drift the proposition describes is, at present, not measured at all. Rigour attaches to the fit between a measure and its object, not to the measure's familiarity.

There is a larger reason, and it is the one that makes the other three cohere. Where the agenda states a direction, it states it in the currencies it can measure: output, consumption, mortality, life-years, public funds. Those are real directions and this paper does not disparage them. But the question the title asks is asked of the transition, and of the institutions deploying AI into it: toward what, for the people living through it? Restrepo models the wage path and then asks about meaning. That is the right order and the right question, and it has no instrument. Growth continues, and, in his title, we won't be missed. That is a direction only if more is the direction.

Wellbeing, measured directly, is the candidate answer, and composition is the steering variable. The first says what a deployment is for, which is what distinguishes a compositional shift that is progress from one that is extraction. The second is what can be observed and moved while the aggregates are still flat. Without the first, the second is a dashboard with no destination. Without the second, the first is a sentiment. Together they are the link between the agenda's mechanics and any institution that has to choose.

A north star gives a direction, not a weighting. Measuring wellbeing does not decide how agency is weighed against security, present gains against future losses, or one group's interests against another's. Those are choices, and they belong to the people and institutions making policy. What measurement does is make them visible as choices, with evidence on each side, rather than settled by default in favour of whatever happened to have a number.

4. THE UNIT OF OBSERVATION: COYLE'S FACTORY COUNT

Coyle and Poquiz's deeper point is about structure rather than scale. AI used as an intermediate input changes

processes as much as it changes inputs and outputs. An input-output table will not show the change; it will eventually show its consequences. Their analogy is the steam engine. Had input-output tables existed in 1800 they would have recorded more steam engines on the input side and more cotton on the output side, and said nothing about the factory system. To see the transformation you would have had to count factories and urban textile workers.

> *That is the observation principle. The transformation is not in the volumes. It is in the composition: the mix of forms through which economic life is organised. The factory count is a compositional measure, and the AI transition needs its equivalent.*

A usable factory count has to do three things. It has to be visible early, while the aggregates are still flat. It has to be comparable across places, because the transformation will not arrive evenly. And it has to connect to an outcome people have reason to care about, which is where Section 3 comes in: the outcome is wellbeing, observed directly, and the factory count is what we expect to move with it.

The agenda reasons in compositional variables throughout. Restrepo's central variable is the labour share, a composition of income. Korinek and Lockwood's second stage is defined by the share of resources absorbed by autonomous systems. Brynjolfsson and Hitzig's evidence is four-firm concentration ratios, a composition of markets, and their model explains why shares shift. Stiglitz and Ventura-Bolet's equilibrium is the share of misinformation in the information pool. Restrepo's bottleneck and supplementary work, and Manning and Aguirre's exposed and adaptive quadrants, are compositions of the workforce. Several of these models explain the mechanism behind their share. What none does, and what this paper proposes, is link those separate mechanisms to a common multidimensional outcome framework, so that a movement in any share can be read against what it does to people.

Composition, as used so far, is an organising concept, not yet a measure. The shares just listed have different denominators, units and dynamics; they do not add up to a partition of anything. Section 9 takes up what it would mean to make composition a measure. The rest of the argument takes the agenda at its word on what moves, and asks what the movement is worth in a currency that does not depend on aggregates the agenda itself says will be late.

5. MECHANISM: THREE EXTRACTION CHANNELS IN THE AGENDA'S OWN MODELS

Three of the agenda's chapters describe, in formal terms, processes by which measured efficiency can rise while something the chapter itself treats as valuable is diminished. Athey and Scott Morton describe their mechanism in the language of extraction. The others do not use the word. We think all three are instances of one pattern, with the caveat, developed at the end of this section, that they operate in different units, need separate evidence, and are conditional results rather than demonstrated outcomes.

5.1. Rent Leakage

Athey and Scott Morton model AI as a priced factor supplied by an upstream that may hold market power. The results do not require the AI to be bad at its job; they require that the upstream sets the price and that the economy adjusts in general equilibrium. In their model with two non-traded sectors and discrete adoption, AI lowers unit costs and raises variety in the adopting sector and displaces unskilled workers into sectors with diminishing returns, so their wage can fall. The upstream supplier then faces an adoption constraint: it can price up to the point at which a firm would rather revert to its pre-AI technology at equilibrium wages. A monopolist typically prices on that boundary. Capping a usage fee shifts rent into the access fee and vice versa. Income can leak abroad. Real wages and aggregate welfare in the adopting country can fall even as the technology works exactly as advertised. Their prescription is pressure on both fee instruments together with policy that supports productive absorption of displaced labour.

Read compositionally: the share of domestic value accruing to a foreign upstream can rise, and the share accruing to domestic labour can fall. Output can rise while the composition of claims on it moves away from the people who live there. This channel is a transfer of rents, denominated in money, and a payment to a foreign supplier is not by itself a rent: the service has a competitive price, and extraction is the part above it.

5.2. Hollowing of the Information Commons

Stiglitz and Ventura-Bolet begin from the oldest result in information economics: information is a public good, so it will be undersupplied without intervention. They identify four channels through which AI and platforms act on the ecosystem. Processing and transmission become more efficient. The producer business model, which depends on visits, is endangered as AI answers questions without

sending anyone to the source. The relative cost of producing misinformation falls. And consumers' ability to screen truth from falsehood changes. Their result is conditional and they keep it so: in the absence of adequate regulation, covering accountability, moderation and intellectual property, the equilibrium quantity of truthful information can decline and the share of misinformation rise, with AI's effects potentially more adverse than platforms' alone.

Read compositionally: the institutions that produce reliable information can shrink as a share of the ecosystem while the efficiency of processing the remaining pool rises. This channel is a degradation of an institution, denominated in information quality, which has no money price.

5.3. Centralisation and the Devaluation of People

Brynjolfsson and Hitzig return to Hayek. Decentralised decisions are valuable for two reasons: local knowledge is inalienable, and the capacity to process information is bounded. AI weakens both. In a property-rights model with an entrepreneur holding an information asset and a headquarters holding a physical one, making the information ass*et al*ienable shifts the efficient ownership regime toward the centre, and relaxing the processing bound shifts it further. They are careful about what this does and does not show. Three countervailing forces are listed: limits to what machines can codify or process, reasons for decentralisation that persist even in an all-AI economy, such as communication latency and incomplete contracts, and legal requirements and preferences for human control. Their response to the possibility that AI empowers the periphery as much as the centre is tentative, resting on emerging evidence that in many workloads a single strong agent outperforms multi-agent arrangements. Their evidence of rising concentration, in retail and in manufacturing, banking, telecommunications and airlines, is descriptive, not an estimate of AI's causal effect. Their political conclusion, that economic power begets political power and that the incentive to invest in human capital may fall, is a concern they raise, not an effect they measure.

Ide and Talamàs (2025) supply distributional detail from a related, peer-reviewed model: autonomous AI favours the most knowledgeable workers and widens labour-income inequality, while non-autonomous AI benefits the less knowledgeable. Ide and Talamàs (2026) extend the analysis across countries, where compute is concentrated in advanced economies. Chatterji, Rock and Talamàs (2026a) add machine-tacit knowledge, what machines know but cannot tell, so that the Nonaka (1994) spiral by which organisations convert tacit knowledge to explicit and back now has a non-human loop in it, a link developed in a preprint (Chatterji, Rock & Talamàs 2026b). Agrawal, Gans and Goldfarb supply the labour-market end, under their cost assumptions: routine knowledge workers are displaced if AI becomes efficient enough, and human geniuses retain an advantage only at the questions furthest from existing knowledge.

Read compositionally: the share of economic life that invests in people and relies on their local judgement may fall, and the share that routes around them rise. This channel would be a loss of capability, denominated in human capital, which is neither money nor information.

5.4. What the Three Channels Share, and What They Do Not

Each channel moves value or capacity from a dispersed base toward a concentrated claimant, and each threatens something the chapter itself treats as load-bearing: the outside options of workers, the supply of truthful information, the human capital of a population. Each is invisible in output and visible in composition. That is the pattern we call extraction.

> *Extraction is a shift in the composition of economic life that moves value or capacity from a dispersed base to a concentrated claimant and erodes, relative to a stated counterfactual and over a stated horizon, the base's capacity to generate future output and wellbeing.*

Two things are deliberate in that definition, and one thing must be said about its status. It does not require aggregate output to rise, because extraction that reduces output is extraction still. It does not require a theory of fairness or desert; it requires the agenda's own variables, read for direction and for whose base is being eroded. And its two clauses have different standing: the transfer is what the models describe; the erosion of future capacity is the empirical hypothesis, and a foreign payment or a redistribution of rents does not on its own establish it.

Nor does the definition make the three channels commensurable. Rents are money. Information quality is not. Human capital is not. The models establish each channel separately and under their own conditions; they do not jointly prove that all three satisfy the definition in a single economy, and a lower wage is a flow, not a measured loss of a stock. A framework that uses these channels has to say, for each, whether it is measuring a

rent, a transfer, a loss of capability or the degradation of an institution; whose base and whose welfare; against what counterfactual; over what horizon.

A companion paper (Ricketts & Beaumont 2026a) develops the same pattern and warns that when AI dissolves a friction the gains can migrate to whoever controls the new substrate. Athey and Scott Morton's upstream pricing is rent migration in that sense. This section's contribution is to show that the three mechanisms are already present, in formal dress, in the work of economists who would not describe themselves as writing about extraction.

6. ENVIRONMENT: TRANSITION PATHS MODELLED, NOT YET OBSERVED IN WELLBEING TERMS

The agenda is not silent on the transition, and it is not only prose. Restrepo's Section 1.2 gives formal transition results: Proposition 7 describes the compute-binding path, along which adjustment is gradual, and Proposition 8 the algorithm-binding path, along which wages can jump. The propositions track automation dates, compute values and wages. Korinek and Lockwood give a roadmap of instruments before and during the transition and argue for proactive institutional adaptation. Athey and Scott Morton discuss the difficulty of reallocating workers and name education, training and care as a research agenda. Agrawal, Gans and Goldfarb motivate their short run with organisational rigidities and adjustment costs. Equilibrium models can and here do represent adjustment over time.

What they represent is the wage path. What no chapter does is place alongside that path any observation of agency, purpose, relationships, health or time use for the people walking it. Charles Jones prices catastrophe, what it is worth to avoid an outcome. The others price output or wages along the way. The years in between have a modelled trajectory for income and no observation of anything else. That is the defensible criticism, and it is also an opportunity: Restrepo's propositions give the measurement programme of Section 9 a concrete economic model to accompany. The question is not whether the transition can be modelled. It is what multidimensional wellbeing does along the paths the model describes, and whether the gradual and discontinuous cases differ in that respect as much as they differ in wages.

The nearest thing to an observed transition variable is Manning and Aguirre's adaptive capacity index, and it is instructive in both directions. The index measures what workers in 356 occupations would bring to a job transition if displacement occurred. It is built from characteristics of people, not of tasks. It finds resilience where exposure measures see only risk: of 37.1 million workers in the top exposure quartile, 26.5 million are in occupations with above-median adaptive capacity. And it finds a concentrated pocket of vulnerability, 6.1 million workers in highly exposed, low-adaptability occupations, concentrated in clerical and administrative roles. They are explicit that exposure is a necessary but not sufficient condition for displacement, and that involuntary job loss, where it occurs, carries costs to earnings, health and mortality that persist for years (Jacobson, LaLonde & Sullivan 1993; Sullivan & von Wachter 2009). What they measure is preparedness for a displacement that may or may not come: a conditioning variable for a transition outcome, observed once. No chapter observes an outcome through one.

The right description of the transition is an environment, and the environment has been characterised before, by people who had to operate in it rather than model it. The military and management literatures call it VUCA: volatile, uncertain, complex, ambiguous. Cascio's revision for a world past mere turbulence is BANI: brittle, anxious, non-linear, incomprehensible (Cascio 2020; Cascio, Johansen & Williams 2025). The four letters are structural, affective, dynamic and epistemic phenomena respectively, with no reason to co-move, so we read them as separate channels and not as a composite. Each has a referent in the agenda.

- **Brittle** (structural). An economy dependent on an imported, concentrated AI stack (Athey and Scott Morton), in a region where compute is concentrated elsewhere (Ide & Talamàs 2026), can appear strong and fail suddenly when a price, a licence or an export control moves.
- **Anxious** (affective). The displacement costs Manning and Aguirre document are the near term. The far term is Restrepo's title. An economy in which human work is, in his term, supplementary is one in which the growth model has answered the question of what people are for, and Restrepo is right to turn, as he does, to meaning.
- **Non-linear** (dynamic). B. Jones's bottlenecks mean constrained acceleration: progress limited by what AI cannot yet do, and released when a bottleneck clears. Restrepo's algorithm-binding path admits wage discontinuities by construction.
- **Incomprehensible** (epistemic). Hadfield and Koh's point is that the objectives of today's AI agents are not known, even to their makers, and that the

evidence on their behaviour is thin. Machine-tacit knowledge, in Chatterji, Rock and Talamàs's term, is what machines know but cannot tell. Whether this is permanent is open; that it is the present condition is not.

A companion paper (Ricketts & Beaumont 2026b) proposes that wellbeing may function as adaptive infrastructure, so that the variables of Section 9 would be not only outcomes of the transition but part of what determines how a population absorbs it. That proposal is preliminary and this paper's argument does not depend on it. What carries over is narrower: the AI transition is a disruption load that falls unevenly; the measurable short-run channel is affect, which the OECD module adopted in 9.2 captures; and the case for observing wellbeing longitudinally, with attention to distribution, is the case for observing the transition at all.

Mullainathan adds a dimension the others do not. If algorithms change how decisions are made, then the quality of human decisions under algorithmic advice is itself a transition variable. His judges example, in which Kleinberg and co-authors estimate that an algorithmic risk score could allow up to a 40% reduction in jailing at constant risk, is an estimate of potential, not a deployed result, and Mullainathan is explicit that it is not an argument for automation. It is better read as evidence of what human judgement systematically misses, and therefore of what it would take to educate judgement rather than replace it. Section 8 returns to this.

7. ECONOMIES THAT IMPORT THEIR INTELLIGENCE

Athey and Scott Morton explicitly consider countries that import AI services from foreign suppliers, and they note competition authorities beyond the United States. We build on that rather than claiming it as new. What we add is a reading of their model from the importing country's side, and two policy layers their chapter does not develop.

For an importing country, the upstream is foreign. That does not put competition enforcement out of reach: in April 2025 Japan's Fair Trade Commission issued a cease-and-desist order against Google LLC over the preinstallation and default-search terms of its distribution and revenue-share agreements with Android device makers and mobile carriers (JFTC 2025). An importer can act against a foreign supplier's conduct in its own market. What it cannot do is set the price of the upstream input in the global market; it can influence the terms and prices at which that input enters its own, through procurement and the conditions of deployment. The first half of Athey and Scott Morton's prescription, pressure on both fees, is therefore partial for an importer; the second half, productive absorption of displaced labour, is entirely within reach; and a third layer sits between them.

A country that does not own the weights can still govern the interface. Access terms, data terms, distribution, procurement, liability, and the conditions under which models are deployed into schools, hospitals, courts and local government are governed where the model meets the user, and that layer is domestic. A working paper (Ricketts 2026a) argues that AI safety and AI political economy are, for most of the world, interface governance problems. Interface governance is broader than the policies that move workers' outside options, though it includes them: data rules, liability and deployment conditions act on the rent, information and human-capital channels of Section 5 at once.

Japan is our case, and we are careful about what is being claimed. The frontier models in wide use in Japanese firms are not Japanese models; they are built and priced elsewhere. In a May 2025 web survey of 4,068 internet users in Japan, among companies using generative AI for work the services most used were ChatGPT (52.1%), Microsoft Copilot (42.3%) and Google Gemini (28.5%), and no Japanese-developed service was among those reported (ICT Research & Consulting 2025). The claim is about deployment, not about domestic research capability, which exists, nor about ownership, procurement, market access or enforcement capacity, each of which is a different thing and would need its own evidence if it were being claimed.

Absorption is where composition becomes a local question. Brynjolfsson and Hitzig consider whether AI might empower the periphery as much as the centre and respond tentatively, on the ground that a single strong agent can outperform multi-agent arrangements in many workloads. That is a claim about compute architecture. The question for a region is different: not whether its agents can out-coordinate the centre's, but whether the centre can take what the region has.

Here we state a hypothesis, and we state it as one. AI can codify local knowledge; Brynjolfsson and Hitzig are right about that. We think it cannot, in the same way, codify local trust: the standing of an institution with the people who use it, the legitimacy that lets a platform be adopted rather than imposed, the relationship that makes a user a member rather than a visit. Knowledge is an information asset and becomes alienable. Trust is relational. Trust as a relationship, information about

trustworthiness, reputation systems, adoption and institutional legitimacy are five different things, and local ownership is a sixth; a test of the hypothesis has to measure them separately. A working paper (Ricketts & Beaumont 2026c) makes the claim its organising one: when the code of a platform costs nothing to reproduce, the scarce input is the user base and the institution that holds its trust. The research question, put properly, is under what conditions existing relationships, legitimacy and community governance sustain local adoption despite the scalability of externally supplied AI.

Tomo is the setting we are closest to, and we say so: it was built at the first author's institution. It is a students-only app for the International College of Liberal Arts in Kōfu, built by a student, Enguun Ariunbold, with a small student team. Its features, finding classmates, a notice board, study rooms, group trips, a student-pinned campus map, a reviewed directory of clubs and services, a signal that one is free for the next two hours, could be reproduced by any of the global platforms its users already carry; the code advantage is nil. What Tomo holds that those platforms do not is borrowed from the institution: membership verified by college email address, a real name on every post, listings reviewed before they appear, initiatives voted on by members, mutual friendships with no follower counts, posts that expire, and a stated purpose of getting users off the app and into real life. Each of those is a candidate trust mechanism. We do not claim that feature reproducibility holds everything else constant: the global platforms still differ in existing networks, convenience, moderation, user experience and promotion, and any adoption Tomo achieves will reflect all of those as well as institutional standing. Tomo is therefore a prospective case, an accessible setting in which the mechanisms of the hypothesis could be studied, not an identification strategy. The app launched in 2026 and no adoption, retention or wellbeing data yet exist. Those are the measurements, and the case is listed in 9.5 for that reason.

A second local case makes the importing position concrete. The Chūō Shinkansen maglev will connect Kōfu to Tokyo on a route, timetable and cost decided elsewhere (JR Central n.d.); what the city can act on is the surrounding municipal policy. That is the importing economy's position with a train in place of a model. A prospective design prepared at the authors' institute (Ricketts & Beaumont 2026d) applies the Anti-Fragile Well-Being framework to that choice with a non-compensatory rule: each of the six wellbeing domains has a capability floor, and a proxy that rises while a domain falls through its floor, land value up as affordability falls, is a fail and not a low score. No probe has yet been run. It is cited only for the design logic that Section 9.3 reuses: floors as constraints, distribution as a requirement, and evidence that would change the decision stated in advance. The city's basic policy for the planned maglev station area states its aim, in its own words, of a Yamanashi-style wellbeing life (Kōfu City 2023), which is the municipal form of the problem this paper is about: a stated objective awaiting an instrument.

8. THE FRONTIER TIER: AN EDUCATIONAL HYPOTHESIS

Agrawal, Gans and Goldfarb's model has an educational reading, and it is conditional. Routine knowledge workers apply known answers to questions near what is known. Geniuses generate new knowledge at a cost rising with distance from the known point. Under their assumption that AI's cost parameter is at least as high as a human genius's, AI geniuses push human geniuses to the far frontier. Routine workers are fully displaced only if AI becomes efficient enough; their Proposition 3 allows routine work to survive otherwise, and their conclusion calls for models of complementary human-AI workflows. What the model motivates is a hypothesis: that the capacity to work at the frontier is the educational investment least likely to be stranded.

Talamàs (2026), in conference slides and so in preliminary form, gives the closer basis for teaching. The efficiency trap, his term, is using AI without learning from it: AI directed at the difficulty that builds judgement and not at the difficulty that merely drains capacity. The question for a classroom is which difficulties are worth facing. A student who uses AI to produce answers near what is known is practising the routine tier. The alternative proposed for firms, a human-machine knowledge spiral in which each teaches the other (Chatterji, Rock & Talamàs 2026b, a preprint), is a description of what a classroom can be, and it is testable in one.

Brynjolfsson and Hitzig supply the stakes as a concern rather than a measurement: if the incentive to invest in human capital falls, an institution that still invests in people is swimming against the current, and the justification for doing so is that it produces people the current cannot carry. That, we think, is what the liberal arts are for in the age of AI: not breadth for its own sake, but the capacity to ask the question furthest from the known point and to recognise an answer when one arrives. A working paper (Ricketts 2026b) develops this from classroom practice. The agenda's models do not

prove it. They make it the hypothesis most worth testing, and Section 9 says how.

Mullainathan's judges belong here. If an algorithm can see what a judge systematically misses, the educational task is to learn what was missed and why, so that judgement is improved rather than outsourced. That is the difference between the knowledge spiral and the efficiency trap at the level of a single decision, and it is the line Mullainathan himself draws when he declines to read his result as a case for automation.

9. MAKING COMPOSITION AND WELLBEING OPERATIONAL

Everything above is argument. The contribution has to be a proposed measure, an observation schedule and an appraisal rule, specified closely enough to be built and run, and this section proposes all three. We propose a dashboard rather than a composite index, because the indicators do not share a denominator and a single number would hide the movements between dimensions that Section 3 says are the point.

9.1. The Factory Count: Compositional Indicators

For each indicator: the entity and geography observed; numerator, denominator, period and source; how overlapping categories and value crossing jurisdictions are handled; and whether the indicator describes an institution or measures an outcome. Candidate indicators, each tied to a channel of Section 5:

1. **Upstream claim share** (rent channel). Payments to non-resident AI suppliers as a share of value added in AI-using sectors, by sector and fee type (access versus usage), from transaction and customs data. This is a service-payment ratio, not a partition of value added and not a rent measure. Identifying the rent component requires a competitive-price benchmark, from cross-supplier or cross-jurisdiction price comparison, and the two are reported separately. Institutional indicator.
2. **Verified-information share** (information channel). The share of the domestic information ecosystem produced by institutions with verification incentives and accountability, measured by output, reach and revenue. "Verified" needs a coding rule (editorial standards, correction practice, accountability to an identifiable publisher) before this can be computed. Institutional indicator.
3. **Human-capital-investing share** (capability channel). The share of employment and of firm expenditure in activities that train and develop people, against activities that route around them; adoption mode (tool versus autonomous agent) by occupation, following Ide and Talamàs (2025). "Training" against "routing around" needs an occupational and expenditure coding rule. Institutional indicator.
4. **Local-institution share.** The share of a region's economic activity running through locally held and locally governed institutions, with ownership and governance recorded separately from trust. The unit of "local economic activity" (establishment, transaction, employment) must be chosen and the denominator fixed. Institutional indicator.
5. **Adaptive capacity.** Manning and Aguirre's index, or a local equivalent, by occupation and region. Conditioning variable.

These are candidates. Table **2**, at the end of Section 9.2, consolidates them with the six wellbeing domains and gives each a provisional operational definition, unit, denominator, data source and observation frequency. The rules are provisional: none has yet been applied, and testing them, with inter-coder agreement reported for indicators 2 to 4, is the first item of the counter-agenda.

9.2. The Outcome: Multidimensional Wellbeing, Observed Repeatedly

The outcome side has to be observed directly, at a frequency that sees the transition, on a sampling frame that sees distribution, with attribution rules that say whose wellbeing is being measured when an agent chose. We build on a published framework rather than proposing one from scratch.

Anti-Fragile Well-Being (Yadava, Achoh, Yamazaki, Beaumont & Ricketts 2025) specifies six wellbeing domains: health, psychological flourishing, social connection, civic engagement, economic security and environmental quality. It pairs conventional indicators in each domain, survey measures and administrative statistics, with a narrative layer drawn from the Virtual Living Lab (VLL), which classifies publicly available online discourse by theme and emotion at scale. The published implementation covers Japan, the United Kingdom, India and Cameroon in four languages; its data, classifiers, validation and ethics protocol are set out in the paper and not repeated here. The framework's purpose is adaptive policy: a five-stage cycle of diagnosing fragility, analysing cultural levers, designing culturally aligned policy probes, monitoring wellbeing and narrative feedback, and refining, replicating or retiring, on

Table 2: Consolidated indicator specification: institutional indicators and wellbeing domains, with provisional operational definitions, units, denominators, data sources and observation frequencies

Indicator	Operational definition (provisional)	Unit and denominator	Data source (Japan as the worked example)	Frequency	Wellbeing domain
Institutional indicators					
1. Upstream claim share	Payments by resident firms and public bodies to non-resident suppliers of AI services (model access, usage, AI-specific cloud), split into access and usage fees. The rent component is reported separately, as the excess over a competitive-price benchmark from cross-supplier or cross-jurisdiction prices.	Per cent. Denominator: gross value added of AI-using sectors.	Balance-of-payments services statistics (computer services; Ministry of Finance and Bank of Japan). These do not separate AI services, so the AI share and the fee split need a firm survey and public procurement records.	Quarterly; annual for the fee split	Economic security
2. Verified-information share	Share of news and public-information output from outlets that meet all three criteria: published editorial standards, a public corrections practice, and an identifiable, legally accountable publisher. Outlets are coded by two independent coders.	Per cent. Denominators, reported separately: audience reach; revenue of the news and information sector.	Audience-measurement panels; industry revenue statistics; coded outlet register.	Annual	Civic engagement
3. Human-capital-investing share	(a) Employer training expenditure and training hours. (b) Adoption mode: share of employees whose main AI use is a tool they direct, against an autonomous agent that performs the task.	(a) Per cent of labour cost; hours per employee per year. (b) Per cent of employment, by occupation.	Employer training survey (Basic Survey of Human Resources Development, Ministry of Health, Labour and Welfare); a labour-force or establishment survey with an added AI-use module.	Annual	Psychological flourishing; economic security
4. Local-institution share	Share of activity in a defined function or region carried by establishments that are locally headquartered and locally governed (a majority of the governing body resident). Ownership and governance are coded separately. Trust is measured in the survey, not inferred from ownership.	Per cent. Denominator: employment in the function or region (primary); transaction value (secondary).	Business register and Economic Census; procurement records.	Annual; quarterly for a single function	Civic engagement; social connection
5. Adaptive capacity	The Manning and Aguirre (2026) index, or a local equivalent built on the same logic, by occupation and region.	Index score; per cent of employed persons in high-exposure, low-capacity occupations.	Occupational and household surveys (Employment Status Survey; Labour Force Survey).	Annual where data allow	Conditioning variable: every domain is reported by band
Wellbeing domains (subjective and objective indicators reported separately)					
6. Health	Subjective: self-rated general health. Objective: healthy life expectancy and mortality for populations; sickness absence for the staff panel.	Score distribution and per cent below floor; years; rate per population; days per employee.	Panel survey; vital statistics; personnel records.	Survey monthly (staff) or quarterly (residents); records annual	Health
7. Psychological flourishing	Subjective: life evaluation, affect and eudaimonia from the OECD (2025) modules; for staff, items on autonomy and purpose at work. No objective counterpart is claimed.	Mean, distribution and per cent below floor, per respondent group.	Panel survey.	Monthly (staff); quarterly (residents)	Psychological flourishing
8. Social connection	Subjective: perceived social support; loneliness. Objective: frequency of contact with family, friends and neighbours.	Score distribution; contacts per week; per cent with no one to rely on.	Panel survey; time-use survey (Survey on Time Use and Leisure Activities) for benchmarks.	Quarterly	Social connection
9. Civic engagement	Subjective: trust in institutions, including the institution deploying the AI service. Objective: voter turnout; participation in local associations; complaints and appeals about the service.	Score distribution; per cent of eligible residents; appeals per 1,000 service contacts.	Panel survey; electoral and administrative records.	Quarterly; records annual	Civic engagement

10. Economic security	Subjective: financial worry; ability to meet an unexpected expense. Objective: household income; labour-income share of household resources; employment status; benefit receipt.	Score distribution; currency per equivalised household; per cent.	Panel survey; tax and benefit records where linkage is lawful and consented.	Quarterly; records annual	Economic security
11. Environmental quality	Subjective: satisfaction with the local environment. Objective: air quality; access to green space. The weakest expected link to an AI deployment; reported for completeness and as a check that movements elsewhere are not general survey drift.	Score distribution; pollutant concentration; per cent of residents within a set distance.	Panel survey; environmental monitoring.	Annual	Environmental quality

Table 2 is a specification, not a result. The Japanese sources are named to show that each row can be populated from existing collections plus a small number of added survey items; the equivalent sources differ by country.

feedback cycles of months rather than years. Its monitoring arrangements are proposed and differ across its cases; the paper is explicit that its policy initiatives are prototypes and that large-scale evaluations are in progress. It does not establish an already implemented person-level panel at a single frequency, and we do not claim one. The observation schedules below are this paper's proposed design.

Three kinds of evidence are kept distinct in what follows, because they answer different questions. *Narrative signals* from the VLL layer are indicators of concern and theme; they are fast and broad and are not a measure of anyone's wellbeing. For them to count as early warning, the design pre-specifies the lead time sought, the predictive improvement over a baseline that would justify the claim, and an acceptable false-alarm rate; absence of earlier movement in one study would not establish that narrative monitoring has no other value. *Measured wellbeing outcomes* are the domain indicators, observed on a defined sample at a defined frequency; for the subjective components we adopt the OECD's (2025) updated survey modules for life evaluation, affect and eudaimonia, so that results are comparable with national and international series. *Causal evidence* that an AI deployment changed an outcome requires a design, a comparison and a counterfactual; neither of the first two kinds supplies it on its own.

Distribution and adaptive monitoring are already concerns of the published framework. What this paper adds is their specification for AI exposure, delegation and institutional arrangements. *Institutional linkage*: each indicator in 9.1 is paired with the domains it is expected to move, so that the dashboard reads composition and outcome side by side (upstream claim share with economic security; verified-information share with civic engagement; human-capital-investing share and adoption mode with psychological flourishing and economic security; local-institution share with civic engagement and social connection). *Delegation attribution*: where an agent acts for a person, outcomes are attributed to the principal and the delegation is recorded as a covariate, following Perrier. Divergence between the agent's objective and the principal's wellbeing is not itself observable, since the objective may be unobserved and the two are in different units; the design therefore records observable proxies, decision discrepancies against resident-endorsed goals, overrides, appeals and subsequent outcome measures, and states which proxy stands for alignment. *Distribution*: every domain is reported by exposure quartile and adaptive-capacity band, following Manning and Aguirre, not only in aggregate. *Transition tracking*: the same individuals are observed repeatedly, staff monthly and residents quarterly in the design of 9.3, so that change along the transition can be seen while it happens.

Subjective and objective indicators. Within a domain the two are reported side by side and are never averaged or converted into one another. They are not interchangeable measures of one quantity. They can diverge, and the divergence is informative: income can rise while financial worry rises, and reported satisfaction can hold through adaptation while an objective condition worsens. The appraisal rule of 9.3 therefore names, for each floor, the indicator on which it is set.

Adaptation across contexts, and inequality within them. The sources in Table **2** assume the administrative data of a high-income economy, and several will not exist elsewhere. The design therefore has a minimum form. The subjective modules and a short set of domain items can be fielded by survey, including by mobile phone, where administrative records are thin, and the institutional indicators can be coded for a single sector or city from procurement records and a purpose-built establishment survey in place of national accounts. The published framework was implemented in Japan, the United Kingdom, India and Cameroon, which is one reason for building on it. Where informal employment is large, labour-income share and occupation-based

adaptive capacity need local redefinition, and indicators are not compared across countries until their comparability has been tested (Section 9.4). Within any population, results are reported as distributions and not only as means: by exposure quartile and adaptive-capacity band, and also by income group, sex, age, occupation, region and urban or rural residence, with the share of each group below the domain floor. Small groups are oversampled so that their estimates are usable, and an aggregate improvement does not satisfy the appraisal rule of 9.3 if a group falls through a floor.

9.3. A Worked Research Design: Two Deployment Arrangements

The design is hypothetical and stated so that it could be run. It demonstrates appraisal; it does not by itself test the Section 3 proposition, which needs the household-resource, free-service-exposure and delegation measures of the panel study in 9.5. A prefectural government is deploying an AI assistant across its resident-services function: enquiries, applications, benefits guidance, referrals. Two arrangements are on the table.

Arrangement A. An imported assistant, procured on usage fees, hosted and operated by the vendor, with resident data processed outside the prefecture. Front-line staff are reduced; remaining staff monitor exceptions. Service quality and unit cost are the vendor's stated metrics.

Arrangement B. The same frontier model, accessed through a locally held institution (the prefecture itself or a local cooperative) under interface terms: data residency, human review of consequential decisions, published escalation rules. Front-line staff are retained and retrained to direct and check the assistant rather than be replaced by it. Unit cost is higher.

What is being compared. A and B differ in staffing, training, oversight, governance and data handling together. The estimate is of the whole deployment arrangement. It cannot isolate the effect of local trust or local ownership, and the design does not claim that it can.

Affected populations. Residents who use the service, stratified by age, income and digital access, together with residents who cannot access it or stop using it, who are followed deliberately because the harms being investigated may show up among them first; the baseline staff cohort, followed after redundancy or transfer as well as in post, by occupation and adaptive-capacity band; and, as a secondary population, local organisations to which the service refers.

Institutional measures (from 9.1): upstream claim share, with service payments and any rent component reported separately; adoption mode by occupation; human-capital-investing share, as training hours and expenditure; local-institution share for the function.

Wellbeing measures (from 9.2): for residents, economic security, civic engagement (including institutional trust) and social connection, quarterly, OECD modules plus domain indicators; for staff, psychological flourishing (including agency and purpose), economic security and health, monthly on a panel; VLL narrative monitoring of discourse about the service and the region, continuous, with the early-warning criteria of 9.2 pre-specified. Delegation recorded wherever the assistant acts on a resident's behalf, with the alignment proxies of 9.2.

Causal design. Rollout is staggered across municipalities within the prefecture. The unit of assignment is the municipality. Where politically feasible, a lottery allocates both the arrangement and the rollout date; where not, arrangement follows municipal choice and the comparison group is formed by matching on baseline composition, service volume and wellbeing. The follow-up is at least eight quarters after full rollout. Two pre-rollout observation waves are collected, which inform but do not establish the parallel-trends assumption. Spillovers between adjacent municipalities and panel attrition, especially among displaced staff and non-users, are measured and reported, not assumed away. The estimator for staggered adoption with heterogeneous effects is the one of Callaway and Sant'Anna (2021), pre-registered, in place of a two-way fixed-effects regression that would be biased under treatment-timing variation.

Selection and sampling. All municipalities in the prefecture are eligible and are assigned as above. Within each municipality, residents are sampled from the resident register, stratified by age, income band and digital access, with oversampling of residents with low digital access and of service users identified from administrative records; non-users stay in the panel. Staff are a census of the baseline resident-services workforce, not a sample. Sample sizes are fixed before rollout by a power calculation on the pre-specified thresholds.

Baseline, follow-up and evaluation of differences. The baseline is the two pre-rollout waves; follow-up is monthly for staff and quarterly for residents to at least eight quarters. The quantity evaluated is the difference

between arrangements in the change from baseline, for each domain and for each exposure and adaptive-capacity band, reported with confidence intervals and read against the pre-specified thresholds, not against statistical significance alone.

Confounding and the limits of attribution. Named confounders are measured and either balanced by design or adjusted for: municipal size, age structure and income; baseline service volume and digital access; concurrent national or prefectural policy changes; local labour-market shocks; and seasonal patterns in affect. Where assignment follows municipal choice, unobserved differences between municipalities that choose A and those that choose B cannot be excluded, and the estimate is reported as an association. Three limits remain even under a lottery. The number of municipalities is small, so estimates will be imprecise. Staff and residents know which arrangement they are in, so reported wellbeing may respond to being studied or to expectations about the arrangement. And a result from one prefecture and one service function may not carry to others. The design can show that one whole arrangement did better than another in this setting. It cannot show which component of the arrangement was responsible, and it does not license a general claim about imported or local AI.

How evidence enters appraisal. Stated in advance, which is the point of Section 3, and structured as the municipal design of Section 7 is: floors are non-compensatory. The rule has three parts. A service-quality standard, defined before rollout, that any arrangement must meet. Wellbeing requirements: resident institutional trust and staff flourishing may not fall through pre-specified thresholds, and no exposure quartile may bear a measured loss on economic security. And a cost objective: among arrangements that meet the quality standard and all wellbeing requirements, the one with the lower total cost over the follow-up horizon is preferred. Everything else is an explicit trade-off, reported with its evidence.

The decision rule. Four cases, with a fifth state for uncertainty. If both arrangements qualify, the cheaper is preferred and the difference is reported as the incremental cost of B relative to A. If only A qualifies, A is preferred. If only B qualifies, B is preferred despite its cost, and the difference is reported as the incremental cost of B relative to A, not as the least possible cost of respecting the requirements, which would require other feasible arrangements to have been considered. If neither qualifies, neither is adopted and the prefecture returns to design with the evidence on which requirement failed. If the evidence is inconclusive within the pre-specified observation period, because confidence intervals straddle a threshold or attrition exceeds its pre-specified bound, the pre-specified response is to extend a bounded pilot or redesign it, and A is not adopted by default merely because a benefit of B was not established. Failure to establish a benefit is not evidence of equivalence or of acceptable harm.

Findings that would change the preferred policy. A can win: if it delivers the quality standard at lower cost and the requirements hold. B can win: if A breaches a requirement and B meets all of them. Neither can win. The trust hypothesis of Section 7 can take a hit, if B's expected wellbeing advantage does not appear over the horizon, though that would speak to the whole arrangement and not to trust alone. The narrative layer can earn or fail to earn its place against its pre-stated criteria. Each is a finding, not a foregone conclusion, and the design is only worth running if each can occur.

The five-stage cycle of the published framework maps onto this directly: the fragility diagnosed is dependence on an imported stack; the cultural lever is the standing of local institutions; the probes are A and B; the monitoring is 9.2; and refine, replicate or retire is the decision the findings inform.

9.4. Validation Strategy

Validation proceeds in four steps, each reported whatever the result. Reliability: internal consistency for multi-item scales, test-retest stability over a short interval in a subsample, and inter-coder agreement for the coded institutional indicators of Table **2**. Comparability across cultures and groups: measurement invariance is tested across languages, countries and the subgroups of 9.2 before any means are compared; where it fails, comparison is restricted to change over time within a group, and anchoring vignettes are used to assess differences in scale use. Convergent and discriminant validity: subjective and objective indicators in the same domain should be related but not identical, and the narrative layer is tested against the survey measures, not assumed to track them. Criterion validity of the institutional indicators: whether a change in an indicator predicts later change in its paired wellbeing domain, out of sample, against a baseline that uses volume measures alone.

Each institutional indicator is a hypothesis about an outcome, not a measure of it. The relationship between

each and wellbeing has to be tested, with alternative explanations, weighting choices and uncertainty assessed. Local ownership, local trust and wellbeing are three variables. We expect them to be related; we do not assume it, and the design has to be able to find that they are not. That is not a formality. It is entirely possible that the volume of what an economy spends on its institutions predicts wellbeing better than the composition of that spending, and a design that could not return that result would not be a test. The composition-first framing of Section 4 is a claim about where transformation becomes visible first, not a claim about what matters most for wellbeing; the second is what Section 9 exists to find out.

9.5. A Counter-Agenda

Questions this reading of the agenda leaves open, each stated so that it could be a paper.

1. **Build the dashboard.** Test and refine the provisional coding rules and denominators of Table **2**, then pilot them for one economy at quarterly frequency from administrative and transaction data, with the rent benchmark in indicator 1 estimated separately. Coyle and Poquiz call for the data; this specifies the use.
2. **Test the proposition.** A panel study on an existing multidimensional dataset, or the one built in item 1. Specify the consumption measure, the target population and controls; define the three exposures (labour-income share of household resources; a free-service exposure indicator, with its denominator chosen and justified, since an expenditure share assigns zero to the services in question and time or usage episodes give different answers; and a record of consequential delegation); and test whether consumption-only predictions of each wellbeing domain lose out-of-sample performance against richer models as exposure rises. A decline in predictive association is reported separately from any causal claim.
3. **Measure rent leakage for importers.** Athey and Scott Morton's framework estimated for Japan, Korea, the UK and the EU, with access and usage channels separated and service payments distinguished from rents.
4. **Observe wellbeing along modelled transition paths.** Extend Manning and Aguirre from preparedness to longitudinal, distribution-sensitive observation of wellbeing, placed alongside Restrepo's compute-binding and algorithm-binding paths: the morbidity of the transition, not only its mortality (C. Jones) or its endpoint.
5. **Welfare attribution under delegation.** Following Perrier, a wellbeing accounting that attributes outcomes to principals when agents choose and specifies which observable discrepancies stand for misalignment.
6. **Does a multidimensional welfare function move the AGI harvest rate?** Korinek and Lockwood optimise discounted human utility; in their logarithmic benchmark the optimal tax fraction depends on the discount factor alone, and more generally on intertemporal substitution and productivity. A multidimensional extension requires a specified welfare function and an account of how AI investment and taxation act on each dimension. Whether, in which direction and by how much the optimum moves is a derivation, not an assertion.
7. **Study the trust hypothesis.** Case studies of local platforms, Tomo among them, against Brynjolfsson and Hitzig's centralisation model: under what conditions do relationships, legitimacy and community governance sustain local adoption, with trust, reputation, ownership, adoption, convenience and promotion measured separately?
8. **Education as the frontier tier.** Pre- and post-course measures of where students sit on the routine-to-frontier spectrum and the tool-to-agent staircase, using the AI-native surveys already run at iCLA, against Agrawal, Gans and Goldfarb's conditions and the human-machine knowledge spiral.
9. **Interface governance for non-owners.** A comparative study of what importing jurisdictions have done at the interface layer, which of Section 5's channels each instrument acts on, and what domestic enforcement against foreign suppliers has and has not achieved.

9.6. Status of Claims and Alternative Explanations

This paper mixes three kinds of statement, and Table **3** separates them: conclusions supported by existing evidence, theoretical interpretation, and propositions that require empirical testing. Nothing in the third group is asserted as a finding.

Four alternative accounts could each undercut the central argument, and the design is built so that it can return any of them. First, consumption may stay informative. Transfers may sustain security, and purpose and standing may be found outside paid work; if so, the proposition of Section 3 fails and consumption-based

Table 3: Status of the paper's main claims

Claim	Status	Basis
Subjective wellbeing can be measured at population scale, with known limitations (adaptation, scale use, evaluative against experiential measures).	Supported by existing evidence and official practice	OECD (2025); national statistical offices
Involuntary job loss carries persistent costs to earnings, health and mortality.	Supported by peer-reviewed evidence	Jacobson, LaLonde and Sullivan (1993); Sullivan and von Wachter (2009)
Exposure to AI is not displacement, and adaptive capacity is unevenly distributed across occupations.	Supported by evidence for the United States; not yet journal-refereed	Manning and Aguirre (2026)
Official statistics will register AI's transformation late, and free use is missing from household consumption.	Argument from measurement practice; conference-volume chapter	Coyle and Poquiz (2026)
The labour share can converge to zero; rents can leak abroad; truthful information can decline; decision rights can centralise.	Conditional model results under their authors' assumptions; not observed outcomes	Restrepo (2026); Athey and Scott Morton (2026); Stiglitz and Ventura-Bolet (2026); Brynjolfsson and Hitzig (2026)
Autonomous and non-autonomous AI have opposite distributional effects.	Peer-reviewed theory; not an empirical estimate	Ide and Talamàs (2025)
The three channels are instances of one pattern, extraction.	The authors' theoretical interpretation	Section 5
Consumption carries less information about non-material wellbeing as labour-income share falls and free-service use and delegation rise.	Proposition requiring empirical test; may be false	Section 3; Section 9.5, item 2
Local trust is less easily codified than local knowledge and can sustain local institutions.	Hypothesis requiring empirical test; may be false	Section 7; Section 9.5, item 7
The composition of institutions, and not only the volume of spending, is associated with wellbeing.	Hypothesis requiring empirical test; may be false	Section 9.4
Capacity to work at the frontier is the educational investment least likely to be stranded.	Hypothesis motivated by a conditional model	Section 8; Section 9.5, item 8
Wellbeing functions as adaptive infrastructure.	Preliminary proposal from the authors' companion paper; nothing here depends on it	Section 6

appraisal is adequate. Second, adaptation. Reported wellbeing can return toward its earlier level after a change in circumstances, one of the known problems noted in Section 3, so a flat series could conceal a real loss or a real gain; this is why affect and eudaimonia are reported alongside life evaluation and why objective indicators are kept separate. Third, centralised or imported provision may raise wellbeing. Lower prices, better service and wider access are plausible, local institutions can be captured, exclusionary or simply worse, and Arrangement A in Section 9.3 can win. Fourth, any association between composition and wellbeing may run in the other direction or reflect a common cause such as regional income, age structure or prior institutional quality; Section 9.4 treats volume against composition as one such rival. The evidence that would most directly challenge this paper is therefore specific: stable out-of-sample performance of consumption across the exposure bands of the proposition, or no difference in wellbeing between arrangements that differ sharply in composition.

9.7. Implications for Policy and Practice

The design has practical uses that do not wait for the full programme, and they are where it connects to social welfare and human development. For a public body procuring an AI service, the appraisal rule of 9.3 can be written into the procurement itself: a service-quality standard, wellbeing floors for residents and staff, a distributional requirement, baseline measurement before rollout and a pre-agreed response to inconclusive evidence. For national statistical offices the proposal is incremental: add the three exposure measures of the proposition (labour-income share of household resources, use of free AI services, and delegation) to existing wellbeing and time-use surveys, so that wellbeing can be

reported by AI exposure. For competition and digital-policy authorities in importing economies, indicator 1 separates service payments from rents and gives interface governance something to be judged by. For employers and educators, adoption mode and training expenditure are observable now. In human-development terms the six domains are read as capabilities with floors: the question put to a deployment is whether it enlarges what people are able to do and to be, and for whom, not only whether it lowers unit cost. That is the practical meaning of making wellbeing operational: the outcome is observed, by group, before the decision is treated as settled.

10. CONCLUSION

The economics of transformative AI, as the NBER agenda presents it, is an achievement. It takes a speculative premise seriously and follows it to conclusions that are uncomfortable and precise. The labour share goes to zero, along paths the agenda can now describe. The tax base erodes. Rents can leak. Truth is undersupplied without regulation. The centre is favoured. Statistics will see it late.

The agenda also knows what its own results leave out, and says so. Health, meaning, purpose, time, the fairness of the path: these are acknowledged in nearly every chapter, written into several, and appraised, where they are appraised, in mortality, life-years or public funds. None is observed through the transition the agenda models, and none is connected to the institutional choices about AI deployment that are being made now. The agenda provides models of growth, distribution, mortality risk and institutional change. This paper has proposed the operational complement: a research design for repeated, distribution-sensitive measurement of multidimensional wellbeing, linked to deployment arrangements and to explicit rules for appraisal, with a published framework to build on, a consolidated indicator specification, and a comparison that could be run and written into procurement. It is a proposal, and its value will be decided by whether it is run. That contribution does not depend on the agenda having omitted anything; it depends on the agenda having left the transition's human outcomes without an instrument, which it has.

Coyle and Poquiz's steam-engine analogy was offered as a warning about statistics. It is better read as an instruction. Count the factories. For the AI transition, the factories are institutions, the count is the composition of economic life, and the reason to count is to find out, while it can still be changed, what the transformation is doing to the people living through it.

AUTHOR CONTRIBUTIONS

Conceptualization, J.R. and C.B.; writing - original draft preparation, J.R.; writing - review and editing, J.R. and C.B. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DISCLAIMER/PUBLISHER'S NOTE

The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

REFERENCES

Agrawal, A., Brynjolfsson, E., & Korinek, A. (Eds.) (2026). The Economics of Transformative AI. National Bureau of Economic Research Conference Report. University of Chicago Press. ISBN 9780226854083. https://www.nber.org/books-and-chapters/economics-transformative-ai

Agrawal, A., Gans, J. S., & Goldfarb, A. (2026). Genius on Demand: The Value of Transformative Artificial Intelligence. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 1, pp. 15-32). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/genius-demand-value-transformative-artificial-intelligence. Also NBER Working Paper 34316 (2025), https://doi.org/10.3386/w34316

Athey, S., & Scott Morton, F. (2026). Artificial Intelligence, Competition, and Welfare. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 4, pp. 95-114). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/artificial-intelligence-competition-and-welfare. Also NBER Working Paper 34444 (2025), https://doi.org/10.3386/w34444

Brynjolfsson, E., & Hitzig, Z. (2026). AI's Use of Knowledge in Society. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The

Economics of Transformative AI (ch. 12, pp. 301-330). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/ais-use-knowledge-society

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. Journal of Econometrics, 225(2), 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cascio, J. (2020). Facing the Age of Chaos. Medium, 29 April.

Cascio, J., Johansen, B., & Williams, A. F. (2025). Navigating the Age of Chaos: A Sense-Making Guide to a BANI World That Doesn't Make Sense. Berrett-Koehler. ISBN 9798890571212.

Chatterji, A., Rock, D., & Talamàs, E. (2026a). Transformative AI and Firms. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 7, pp. 173-188). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/transformative-ai-and-firms

Chatterji, A., Rock, D., & Talamàs, E. (2026b). The Human-Machine Knowledge Spiral. arXiv:2606.29227. Preprint, not peer reviewed.

Coyle, D., & Poquiz, J. L. (2026). Making AI Count: The Next Measurement Frontier. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 3, pp. 67-87). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/making-ai-count-next-measurement-frontier. Also NBER Working Paper 34330 (2025), https://doi.org/10.3386/w34330

Hadfield, G. K., & Koh, A. (2026). An Economy of AI Agents. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 5, pp. 119-138). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/economy-ai-agents. Preprint: arXiv:2509.01063.

Hayek, F. A. (1945). The Use of Knowledge in Society. American Economic Review, 35(4), 519-530.

ICT Research & Consulting (2025). 2025年7月 法人向け生成AIサービス利用動向調査 [Survey of corporate generative AI service use, July 2025]. ICT 総研, 2 July. https://ictr.co.jp/report/20250702.html

Ide, E., & Talamàs, E. (2025). Artificial Intelligence in the Knowledge Economy. Journal of Political Economy, 133(12), 3762-3800. https://doi.org/10.1086/737233

Ide, E., & Talamàs, E. (2026). The Impact of AI on Global Knowledge Work. Journal of Monetary Economics, 157, 103876. https://doi.org/10.1016/j.jmoneco.2025.103876

Jacobson, L. S., LaLonde, R. J., & Sullivan, D. G. (1993). Earnings Losses of Displaced Workers. American Economic Review, 83(4), 685-709.

Japan Fair Trade Commission (2025). JFTC Issues a Cease and Desist Order to Google LLC. Press release, 15 April. https://www.jftc.go.jp/en/pressreleases/yearly-2025/April/250415.html

Jones, B. F. (2026). Artificial Intelligence in Research and Development. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 2, pp. 37-61). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/artificial-intelligence-research-and-development. Also NBER Working Paper 34312 (2025), https://doi.org/10.3386/w34312

Jones, C. I. (2026). How Much Should We Spend to Reduce AI's Existential Risk? In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 16, pp. 431-449). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/how-much-should-we-spend-reduce-ais-existential-risk. Also NBER Working Paper 33602 (2025), https://doi.org/10.3386/w33602

JR Central (n.d.). Yamanashi and the planned maglev connection. https://market.jr-central.co.jp/lineartown/yamanashi/1

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human Decisions and Machine Predictions. Quarterly Journal of Economics, 133(1), 237-293. https://doi.org/10.1093/qje/qjx032

Kōfu City (2023). (仮称)リニア山梨県駅前エリアのまちづくり基本方針 [Provisional basic policy for town development in the maglev Yamanashi station-front area]. November. https://www.city.kofu.yamanashi.jp/rinia/taisaku/machikankyou/kotsuu/documents/kihonhoushin_kakutei.pdf

Korinek, A., & Lockwood, L. M. (2026). Public Finance in the Age of AI: A Primer. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 15, pp. 397-425). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/public-finance-age-ai-primer. Also NBER Working Paper 34873 (2026), https://doi.org/10.3386/w34873

Manning, S. J., & Aguirre, T. (2026). How Adaptable Are American Workers to AI-Induced Job Displacement? In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 8, pp. 197-218). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/how-adaptable-are-american-workers-ai-induced-job-displacement. Also NBER Working Paper 34705 (2026), https://doi.org/10.3386/w34705

Mullainathan, S. (2025). Economics in the Age of Algorithms. AEA Papers and Proceedings, 115, 1-23. https://doi.org/10.1257/pandp.20251118

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization Science, 5(1), 14-37. https://doi.org/10.1287/orsc.5.1.14

OECD (2025). OECD Guidelines on Measuring Subjective Well-being (2025 Update). OECD Publishing, Paris. https://doi.org/10.1787/9203632a-en

Perrier, E. (2026). Post-AGI Economies: Autonomy and the First Fundamental Theorem of Welfare Economics. arXiv:2604.21216. Preprint, not peer reviewed.

Restrepo, P. (2026). We Won't Be Missed: Work and Growth in the AGI World. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 9, pp. 225-242). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/we-wont-be-missed-work-and-growth-agi-world. Also NBER Working Paper 34423 (2025), https://doi.org/10.3386/w34423

Ricketts, J. (2026a). Beyond Model Weights: AI Safety as Interface Governance and Political Economy. Working paper, AI+Wellbeing Institute. Not peer reviewed.

Ricketts, J. (2026b). Slow Slop, Fast Slop: Insight Is What the Liberal Arts Must Build in the Age of AI. Working paper, AI+Wellbeing Institute. Not peer reviewed.

Ricketts, J., & Beaumont, C. D. (2026a). Dissolve What Extracts: AI, Wellbeing, and the Composition of Economic Life. A compositional theory of the AI wellbeing dividend, with a constructive policy triad. Wellbeing Futures: Innovation, Policy and Research, 2, 47-60. https://doi.org/10.65638/2978-882X.2026.2.05.

Ricketts, J., & Beaumont, C. D. (2026b). Wellbeing, VUCA(t), and BANI: The Unequal Human Burden of Global Disruption. A formative-vector framework and country-year panel evidence, 2005-2022. Wellbeing Futures: Innovation, Policy and Research, 2, 61-75. https://doi.org/10.65638/2978-882X.2026.2.06

Ricketts, J., & Beaumont, C. D. (2026c). Code Is Free, Users Aren't: AI, Institutions and Local Platforms. Working paper, AI+Wellbeing Institute. Not peer reviewed.

Ricketts, J., & Beaumont, C. D. (2026d). From Maglev to Meaningful Place: Anti-Fragile Wellbeing Design for Kōfu. A Prospective Policy-Design Study of Municipal Choices Ahead of the Chūō Shinkansen. Wellbeing Futures: Innovation, Policy and Research, 2, 76-97. https://doi.org/10.65638/2978-882X.2026.2.07

Shahidi, P., Rusak, G., Manning, B. S., Fradkin, A., & Horton, J. J. (2026). The Coasean Singularity? Demand, Supply, and Market Design with AI Agents. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 6, pp. 145-165). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/coasean-singularity-demand-supply-and-market-design-ai-agents. Also NBER Working Paper 34468 (2025), https://doi.org/10.3386/w34468

Stiglitz, J. E., & Ventura-Bolet, M. (2026). The Impact of AI and Digital Platforms on the Information Ecosystem. In A. Agrawal, E. Brynjolfsson, & A. Korinek (Eds.), The Economics of Transformative AI (ch. 14, pp. 367-389). University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/impact-ai-and-digital-platforms-information-ecosystem. Also NBER Working Paper 34318 (2025), https://doi.org/10.3386/w34318

Sullivan, D., & von Wachter, T. (2009). Job Displacement and Mortality: An Analysis Using Administrative Data. Quarterly Journal of Economics, 124(3), 1265-1306. https://doi.org/10.1162/qjec.2009.124.3.1265

Talamàs, E. (2026). Transformative AI and Firms. Slides, AI in Action Conference: Making AI Work, World Bank, January. Based on work with E. Ide, A. Chatterji and D. Rock. Conference slides, not peer reviewed. https://thedocs.worldbank.org/en/doc/3be18bca77301cb63a54067d8477e758-0070062026/original/9-Talamas.pdf

Yadava, M. K., Achoh, S., Yamazaki, K., Beaumont, C. D., & Ricketts, J. (2025). Anti-Fragile Well-Being: A Cultural Systems Framework for Adaptive Public Policy. Wellbeing Futures: Innovation, Policy and Research, 1, 53-71. https://doi.org/10.65638/2978-882X.2025.01.05

©2026 Ricketts and Beaumont. Published by Wellbeing Futures: Innovation, Policy and Research. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited. (http://creativecommons.org/licenses/by-nc/4.0/)