The Broken Job Ladder:

AI, Vanishing Apprenticeships, and the Role of Universities

Miao Ben Zhang*

September 2026

Abstract

Generative AI disproportionately displaces entry-level cognitive jobs that historically served as apprenticeships through which college graduates developed expert judgment. I build a dynamic model in which AI breaks that ladder by making junior work uneconomical and eliminating learning-by-doing—previously a byproduct of junior production. Employee poaching and managerial myopia then lead firms to underhire juniors, resulting in long-term depletion of senior expertise. Standard firm-side corrections fail under myopia. Universities, by designing micro-residency programs delivering immediate value for firms while embedding judgment development, achieve dominant-strategy incentive compatibility—identifying universities as best positioned to address this structural crisis in human capital formation.

JEL: D86, G30, I23, J24, J31, M53, O33

Keywords: AI, human capital, job ladders, liquidity constraints, managerial myopia, apprenticeship, university education

*Miao Ben Zhang is from the University of Southern California Marshall School of Business and can be reached at miao.zhang@marshall.usc.edu. I thank Nan Jia, Song Ma, Paul Oyer, and Raman Randhawa for helpful conversation and comments that substantially improved the paper. All remaining errors are my own.

Most skilled professions share a common developmental architecture. Universities teach *knowledge*—codifiable, transferable principles such as legal doctrine, accounting standards, and programming languages. Firms then develop *judgment*—the tacit, experience-based capacity to act decisively when the problem is undefined, the stakes are irreversible, and the right answer depends on reading people rather than analyzing documents.[1] Judgment is the product of thousands of supervised decisions accumulated over years of practice. A novice enters the workforce with knowledge, performs routine tasks under supervision, and gradually develops the judgment that distinguishes an expert from a beginner. For decades, production and judgment formation were bundled: the same activity that generated output for the firm also developed expertise in the worker.

Motivated by evidence that generative AI is disproportionately displacing entry-level professional employment,[2] this paper asks what happens to the pipeline through which novices become experts when the routine work that once trained them can be bought from a machine. The question is not whether AI eliminates or creates jobs—every wave of automation has done both[3]—but whether the market, left to itself, continues to finance the human capital that AI cannot produce.

I study this question in a dynamic model in which AI is a production technology and nothing more. Each firm must complete a fixed volume of routine cognitive work, which can be done by junior employees or purchased from a machine at a price that falls when generative AI arrives. Senior professionals, who exercise judgment, are produced only by juniors who have done routine work under supervision. Before AI the machine is unavailable, juniors are essential to production, and judgment forms as a costless byproduct in the sense of Arrow (1962): there is no training decision to make, and the frictions that dominate the training literature—the poaching externality of Becker (1964) and managerial myopia—are dormant. AI changes one number, the price of a unit of routine work. Once that price falls below what a junior costs, employing a junior carries a premium over the machine, and the premium is the price of training. The consequences for junior hiring, junior and senior wages, and the stock of seniors are then equilibrium outcomes. Figure 1 summarizes the argument.

[1] For example, judgment enables a senior M&A partner to walk away from a deal whose financials are clean because the management team's behavior signals undisclosed risk.

[2] See, for example, Brynjolfsson, Chandar, and Chen (2025), Hosseini and Lichtinger (2025), Eloundou et al. (2024), Noy and Zhang (2023), among others. Section 1 details the evidence.

[3] The task-based framework of Acemoglu and Autor (2011), Acemoglu and Restrepo (2019), Acemoglu and Restrepo (2022), and Eisfeldt, Schubert, and Zhang (2023) provides the canonical treatment of how automation displaces workers from tasks. Autor et al. (2024) and Autor et al. (2026) extend this framework to study the creation of new work. This paper differs in that it focuses not on the displacement of tasks but on the financing of the *developmental function* that task performance served.

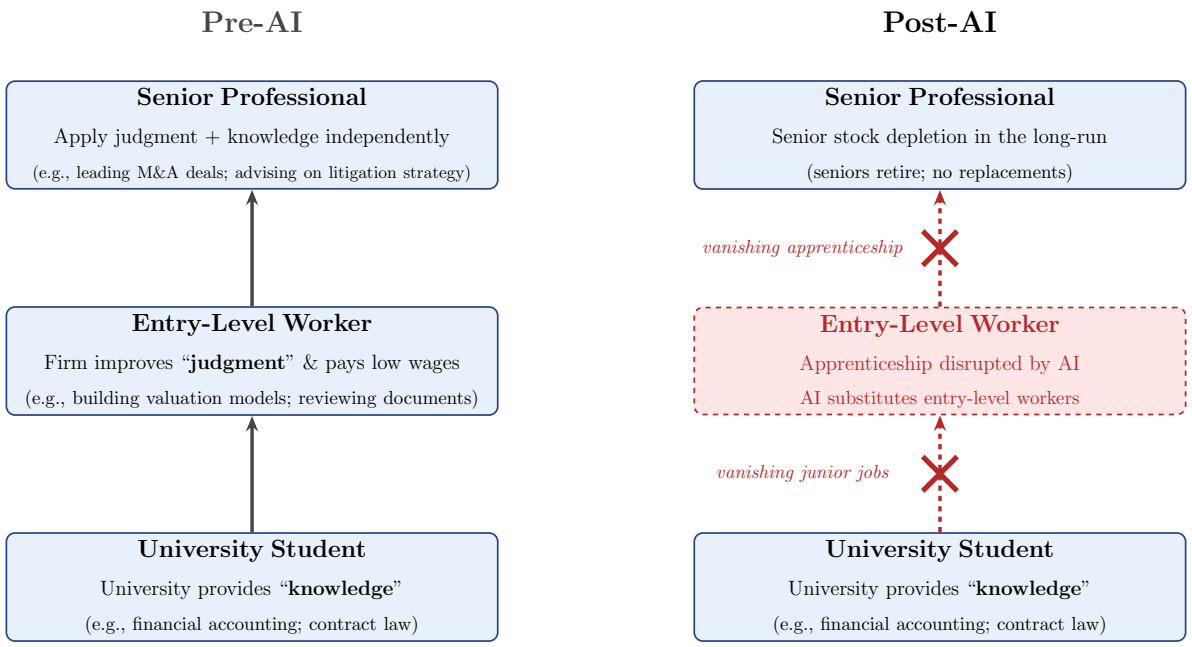


Figure 1: **AI and the Job Ladder.** The left panel illustrates the pre-AI career ladder in which university education (knowledge) and firm apprenticeship (judgment) combine to produce senior professionals, with juniors essential to production and learning a costless byproduct. The right panel shows the post-AI economy: the machine does routine work more cheaply, so a junior costs a premium that someone must finance. When a liquidity floor prevents workers from financing it, the pipeline thins and the senior wage premium rises.

The first result is that, with flexible wages, the ladder largely survives. The value of becoming a senior accrues to the worker, and as AI thins the pipeline the senior wage rises, so workers are willing to accept lower junior wages in exchange for the training—the classical result of Becker (1964) that workers finance general human capital. When workers can pay, neither poaching nor managerial myopia does much damage: if the firm undervalues the future, the worker, who does not, absorbs the difference through the wage. The extinction of senior expertise cannot occur with flexible wages, because the price of the scarce input rises and the pipeline is self-limiting.

The second result is that a floor on entry compensation converts this outcome into a market failure. The junior wage that would clear the market after AI is the worker's outside option less her willingness to pay for training; because that willingness rises with the senior premium, the clearing wage falls as the pipeline thins. Graduates may be unable to accept it. The floor is set by living costs net of the graduate's wealth and the credit she can raise against future earnings, together with legal limits on sub-minimum and unpaid work, and it need only exceed the price of the machine's output, which for routine cognitive work is

close to zero. The constraint is one of liquidity and not a prohibition on negative wages: in the illustration the flexible-market wage is positive but still below what a graduate needs to live on. Because private lenders do not accept human capital as collateral, this is the credit constraint that Becker (1964) identified as the reason workers underinvest in general training. When the floor binds, the firm must finance training it cannot sell to the worker, and two frictions activate: each firm captures only the fraction of the return that survives poaching, and a manager with a short horizon values that return at a discount (Stein, 1989). The steady-state senior stock falls below its efficient level and the senior wage premium rises above it. Because the premium rises, the pipeline does not vanish—for any positive horizon firms keep training—but it can shrink enough that the decentralized response to AI leaves the economy worse off than if AI had never arrived, even though the first-best response is to adopt AI and run a thinner pipeline. The comparative statics are those of a liquidity constraint: junior hiring falls further where entry-level workers have less wealth and less access to credit.

These frictions shape the policy menu. Where the binding constraint is liquidity, extending credit to conventional apprenticeships restores the flexible-compensation allocation at no resource cost. Where credit cannot be extended, verified employer subsidies can implement the efficient stock: a payment at the training date is effective whatever the manager's horizon, while a payment deferred to qualification must grow without bound as the horizon shortens. Both require that the developmental content of a position be verified, and a payment for hiring alone subsidizes an observable input rather than the training it is meant to buy. Verification, supervision, and public funds all have costs, so the relevant comparison is among financing arrangements evaluated with the same accounting.

The third result concerns a training intermediary, which I interpret as a university residency. The intermediary enrolls graduates, assigns them the routine work of sponsoring firms under supervision, charges the firm for the deliverable and the graduate a net tuition, and bears a supervision cost. Its advantages are a price it can charge the patient party—tuition is not a wage, is not subject to the floor, and is financed by student credit—and credible certification of training through accreditation and supervised process records, whose cost is part of the supervision cost. It needs no longer horizon than a firm; the horizon in the mechanism is the graduate's. The residency raises the senior stock if and only if the tuition graduates would pay at the market steady state covers the net cost of supervision, student credit covers living costs and net tuition, and the program's capacity exceeds the market's replacement flow; a smaller program displaces firm training one for one without raising expertise. It is a second-best institution. It cannot collect the rents that accrue to whichever firm employs the seniors it produces, so it stops short of the efficient stock unless subsidized;

supervision is a real resource cost, so even the subsidy that implements the original efficient stock does not restore first-best welfare; and because entry is privately profitable but crowds out firm hiring, a residency can be active yet welfare-reducing. Where conventional credit can be extended, that reform does better. The residency's role is where enforceability limits conventional credit but not educational credit, or where public funds are costly, since an unsubsidized residency raises the stock at no fiscal cost.

The model is, at bottom, about the financing of a long-lived intangible asset. Two financial frictions do the work: a liquidity constraint that prevents young workers from financing their own human capital, and a governance friction that shortens the horizon over which firms value assets that take years to mature (Stein, 1989; Graham, Harvey, and Rajgopal, 2005). Human capital that resides in employees rather than in the firm is the canonical asset that corporate finance struggles to value and to finance (Zingales, 2000; Eisfeldt and Papanikolaou, 2013), and AI raises the stakes by removing the mechanism that financed it automatically. The model yields predictions that can be taken to data: within AI-exposed occupations, junior hiring falls and the senior–junior wage gap widens, more so where entry-level workers have less wealth and less access to credit; and firms with longer-horizon ownership should cut developmental hiring less.

The model abstracts from several considerations. Seniors are assumed to remain necessary, because judgment is forged in undocumented, real-time experience that large language models cannot learn from, so AI does not displace the top of the ladder. Firms are symmetric. Training is a one-period spell that converts a junior into a senior with probability m, rather than a persistent process in which unsuccessful juniors remain in training. The senior market is a reduced-form rationed-matching environment with a compressed wage rather than a full matching model. Screening by universities is outside the model. Relaxing these assumptions would change magnitudes, and relaxing some would change the mechanism: with a fully competitive senior market, for instance, the firm's stake in general training disappears and the Becker benchmark is the whole story, and with persistent training the worker's and firm's continuation values must be solved jointly. I do not claim that the logic survives every relaxation, only that the assumptions are the ones the literature on training has typically made.

The paper contributes to two strands of literature. First, it builds on classical human capital theory—Becker (1964), Ben-Porath (1967), Arrow (1962), Acemoglu and Pischke (1998, 1999)—to study the AI-driven destruction of apprenticeship. The prescription that emerges is neither that firms should internalize the training externality nor that universities are uniquely patient, but that the binding constraint is the worker's inability to pay

for general training through the wage, and that the institution which relaxes it is one that can charge tuition.[4] Second, it contributes to the fast expanding literature on AI and labor markets, including empirical discovery of AI's biased impact on different jobs (Brynjolfsson, Chandar, and Chen, 2025; Hosseini and Lichtinger, 2025; Eloundou et al., 2024), task-level productivity effects of generative AI (Brynjolfsson, Li, and Raymond, 2025; Noy and Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2023), macroeconomic frameworks for assessing AI's aggregate impact (Acemoglu, 2024; Jones, 2026; Aghion, Jones, and Jones, 2018), the microeconomics of AI as prediction technology (Gans, 2025a,b), AI's erosion of human learning incentives and long-run knowledge accumulation (Acemoglu, Kong, and Ozdaglar, 2026), and the case for steering AI development toward pro-worker applications (Acemoglu, Autor, and Johnson, 2026; Acemoglu and Restrepo, 2019, 2022; Autor et al., 2024). The paper provides a framework explaining why entry-level displacement differs structurally from previous automation waves and what institutional responses are and are not incentive-compatible, joining recent work on incorporating frontier technology into college education (Biasi and Ma, 2022, 2026).

Section 1 reviews the evidence. Section 2 sets out the model and derives the AI shock. Section 3 characterizes the planner's benchmark and the equilibria with flexible wages, with a liquidity floor, and with managerial myopia, and compares welfare with and without AI. Section 4 studies credit and transfers aimed at firms. Section 5 introduces the training intermediary, its capacity and cost limits, the role of public funds, and a numerical illustration. Section 6 draws implications for university education, and Section 7 concludes. Proofs and numerical derivations are in the Appendix.

1. Motivating Evidence on AI Disrupting Entry-Level Jobs

A rapidly growing body of research documents the impact of generative AI on professional work. I organize the evidence in three categories.

Labor-market displacement. Brynjolfsson, Chandar, and Chen (2025), using ADP payroll records covering millions of U.S. workers, document a 13% relative decline in employment for workers aged 22–25 in AI-exposed occupations since late 2022, while employment for ex-

[4]See Mincer (1974) for the foundational empirical treatment of on-the-job training and Heckman (1998) for a review of returns to human capital investment across the life cycle. Loewenstein and Spletzer (1998) document that employers bear explicit costs of training; Barron, Berger, and Black (1999) provide evidence that training is associated with lower starting wages and faster subsequent productivity growth; Gao, Wang, and Wu (2024) show that finance workers accept low initial compensation in exchange for larger later salary jumps.

perienced workers remained stable. The decline is concentrated in occupations where AI *automates* rather than *augments* tasks—precisely the routine cognitive work that constitutes the apprenticeship substrate in the model. Hosseini and Lichtinger (2025), analyzing 62 million résumés across 285,000 U.S. firms, report that AI-adopting firms reduced junior hiring by 9–10% within six quarters, driven entirely by reduced hiring rather than increased separations. Critically, senior hiring was unaffected, suggesting AI displaces the *entry point* of the career ladder. Eloundou et al. (2024) estimate that over 46% of U.S. jobs could be substantially affected by LLM-complementary software, with highest exposure in white-collar, cognitively routine occupations.

Task-level productivity. Controlled studies consistently show that generative AI substantially raises productivity on routine cognitive tasks. Noy and Zhang (2023) find that ChatGPT access reduces task completion time by 40% and raises output quality by 18% on mid-level writing tasks, with largest gains for lower-ability workers. Brynjolfsson, Li, and Raymond (2025) find a 15% average productivity increase from a generative AI assistant across 5,172 customer-support agents, with 30% improvement for novices but minimal impact on experienced agents. Peng et al. (2023) report that software engineers using GitHub Copilot complete coding tasks 55% faster. That AI augments novices most does not conflict with the displacement evidence: an augmented junior is still redundant when the machine alone completes the task, and the relevant comparison for the firm is the cost of a junior against the cost of the machine, not the junior's productivity with the machine against without.[5]

Organizational consequences. Fuller, Sigelman, and Fenlon (2025) show that firms are reshaping from pyramidal to diamond structures as entry-level tasks are absorbed by AI. Edmondson and Chamorro-Premuzic (2025) articulate the developmental concern at the heart of this paper—that removing junior roles eliminates the apprenticeship through which novices develop expert judgment. Walther (2025) highlights the convergence of AI adoption with senior retirements, creating a pipeline crisis from both ends.

The pattern the model formalizes is that generative AI displaces the routine cognitive work that historically served as the apprenticeship, and firms respond by reducing junior hiring rather than by financing alternative developmental pathways. The model adds the equilibrium consequence the evidence does not yet speak to: what happens to senior wages,

[5] Dell'Acqua et al. (2023) document that AI improved performance on tasks within the "technological frontier" but worsened outcomes on tasks requiring judgment beyond AI's capabilities, consistent with the knowledge–judgment distinction on which the model rests.

and to the willingness of the next cohort to pay for training, as the pipeline thins.

2. The Model

The model is built so that AI enters only as a production technology. I first describe the environment, then the economy before AI, and then derive the consequences of AI's arrival from the firm's problem, at a steady state and along the transition. The maintained parameter restrictions are $0 \leq c < \omega$, $0 < \beta^M \leq \beta < 1$, and $\theta \in (0, 1)$.

2.1 Environment

Time is discrete and infinite, and a period is the length of a junior spell. There is a continuum of identical firms of total mass $N > 0$, indexed by i, each negligible relative to the industry, and each period a continuum of graduates larger than $N\bar{J}$ enters the labor market. Each firm takes the industry wage paths and its matching inflow as given, because an individual deviation cannot change them. Graduates discount at β. Each firm's decision-maker discounts at $\beta^M \leq \beta$, a reduced form for the governance frictions discussed in Section 3.4; I carry β^M throughout so that each result is stated once, and specialize to $\beta^M = \beta$ where indicated.

Routine work and the machine. Each firm must complete R units of routine cognitive work per period—due diligence on the deals it advises, review of the documents it produces, the first draft of every model.[6] A junior employee completes y^J units per period and is paid w_t^J. Once generative AI is available, the machine completes y^{AI} units per unit purchased at price c^{AI}. Let $\bar{J} \equiv R/y^J$ denote the number of juniors that would cover the requirement, and let

$$c \;\equiv\; c^{\text{AI}}\,\frac{y^J}{y^{\text{AI}}} \tag{1}$$

denote the cost of buying one junior's output from the machine, with $c < \omega$ (the machine is cheaper than a graduate's time). Before AI, $c = \infty$; the arrival of generative AI is a fall in c to a finite value. A firm that employs $J \leq \bar{J}$ juniors and buys the remainder of its routine

[6] Fixing the routine requirement bounds the firm's demand for routine work and rules out the unbounded profits that a linear technology would otherwise generate. The constant profit from having the machine complete the routine work is omitted from the profit function below because it is common to all regimes; a decreasing-returns technology in routine output delivers the same results with more notation.

work from the machine spends $c\bar{J} + (w_t^J - c)J$ on routine work. The quantity

$$p_t \equiv w_t^J - c \tag{2}$$

is the premium the firm pays to have a junior rather than the machine do a junior's work. Its size relative to the value of the senior the junior may become is the model's central object.

Seniors. Each firm employs a stock $S_{i,t}$ of senior professionals who produce $G(S_{i,t})$, with $G' > 0$, $G'' < 0$, $\lim_{S\to 0} G'(S) = \infty$, and $\lim_{S\to\infty} G'(S) = 0$. Seniors retire with probability $\delta \in (0,1)$ each period; retirement means leaving the profession for the outside activity, not losing all earnings. Newly qualified seniors begin producing in the next period, and only seniors already present during production face that period's retirement draw. Two features of the senior market matter, and I state them as an assumption because they are not derived from a hiring problem.

Assumption 1 (Senior market)**.** (a) The senior wage is $w_t^S = \theta G'(S_t)$, where S_t is the symmetric stock and $\theta \in (0,1)$. (b) Newly qualified seniors are allocated by an exogenous matching protocol: a fraction $1-\lambda$, with $\lambda \in [0,1)$, of a firm's newly qualified juniors remains with it, and the rest move for non-wage reasons and are assigned uniformly across other firms. Established matches are not traded and persist until retirement. A firm can neither bid seniors away from other firms nor buy additional seniors on demand at the quoted wage, and it cannot change the matching probabilities by bidding. (c) Seniors prefer the profession to the outside option: $\theta G'(S^{\text{pre}}) \geq \omega$, where S^{pre} is the pre-AI stock defined below.

Parts (a) and (b) are a reduced form for bargaining under search frictions, not an unrestricted competitive spot market for seniors. The share $1-\theta$ of a senior's marginal product that the firm retains is the wage compression that Acemoglu and Pischke (1999) show is necessary for firms to have any stake in general training; a fully competitive senior market, in which firms bid freely and earn no rent, is the limiting case $\theta \to 1$. Because $w_t^S < G'(S_t)$ at the symmetric stock, every firm would like more seniors at the quoted wage; part (b) is what prevents it from buying them, and the only channel through which a firm augments its stock is training. The separation probability λ reflects that professional judgment is largely transferable across firms within a profession: Neal (1995) finds that human capital is predominantly industry-specific rather than firm-specific, and Gathmann and Schönberg (2010) show that skills are portable across occupations with similar task content. Since G' rises as S falls, part (c) implies $\theta G'(S) \geq \omega$ on every path on which the stock does not exceed S^{pre}. Two values recur. Let $\sigma_t \equiv (1-\theta)G'(S_t)$ denote the firm's rent on a marginal senior, and

define, for bounded sequences,

$$D_t \equiv w_t^S - \omega + \beta(1-\delta)\, D_{t+1}, \qquad Q_t \equiv \sigma_t + \beta^M(1-\delta)\, Q_{t+1}. \tag{3}$$

D_t is the present value to a worker, at the start of a senior career in period t, of the senior premium over the outside option; Q_t is the present value of the rent on a senior to the firm's decision-maker. At a constant stock S they reduce to the stationary values

$$\Delta(S) = \frac{\theta G'(S) - \omega}{1 - \beta(1-\delta)}, \qquad v^M(S) = \frac{(1-\theta)G'(S)}{1 - \beta^M(1-\delta)}, \tag{4}$$

and I write $v(S)$ for $v^M(S)$ at $\beta^M = \beta$. The social value of a matured senior relative to the outside option is $\Delta(S) + v(S) = [G'(S) - \omega]/[1 - \beta(1-\delta)]$. Both Δ and v^M are decreasing in S: seniors are paid more, and are worth more to firms, when they are scarce. This is the price response that disciplines the pipeline.

Workers, maturation, and the junior wage. A graduate can accept a junior position, enter a residency when one exists (Section 5), or take an outside option worth ω per period, available in every period. A junior who personally performs the routine work for one period qualifies as a senior at the start of the next with probability $m \in (0,1)$ and otherwise returns to the outside option permanently; nonparticipants also remain in it. All employment quantities are continuum masses, so idiosyncratic qualification, mobility, and retirement draws average out within a firm. The parameter m is a reduced form for learning-by-doing in the sense of Arrow (1962), summarizing what one junior spell contributes to a career; it is shaped by the graduate's preparation, which is where the university's traditional knowledge function enters as a complement to on-the-job learning (Cunha and Heckman, 2007), but I treat it as a parameter.[7] A junior trained in period t enters the senior market in period $t+1$, so a graduate accepts a junior position if and only if $w_t^J + \beta m D_{t+1} \geq \omega$. With graduates in excess supply, competition among them drives the junior wage to $\omega - \beta m D_{t+1}$: workers buy training with a markdown, as in Becker (1964) and Acemoglu and Pischke (1998). The model's central friction is that the markdown is bounded.

Assumption 2 (Liquidity floor and credit)**.** Workers have linear utility over consumption above a subsistence requirement $\ell \in [0, \omega)$, and each entrant holds liquid assets $a \geq 0$. A conventional employee can borrow at most $b_F \geq 0$ at entry; a student in a qualifying

[7] Because unsuccessful juniors leave, $1/m$ is not an apprenticeship duration. A persistent-training variant, in which unsuccessful juniors stay and a junior stock evolves, changes the worker's and the firm's continuation values and is left to future work.

accredited program can borrow at most $b_U \geq b_F$. Loans are fairly priced and fully recourse, with no resource cost or default subsidy: a principal d at entry is repaid by the constant annuity $(1-\beta)d/\beta$ from the next period on, whose entry-date present value is d, and

$$\frac{1-\beta}{\beta}\, b_U \;\leq\; \omega - \ell. \tag{5}$$

A legal minimum wage $w_{\min} \in [0, \omega]$ applies to conventional employees. Every conventional job therefore pays at least

$$\underline{w} \;\equiv\; \max\{\, w_{\min},\; \ell - a - b_F \,\}, \qquad 0 \leq \underline{w} \leq \omega.$$

The limits b_F and b_U are primitives that reflect differences in enforceability and loan eligibility; the model does not assume that a program that calls its charge tuition automatically qualifies for the higher limit. Condition (5) makes repayment feasible even for unsuccessful trainees, who earn at least ω in later periods. Debt service is due whether the worker enters the profession or the outside activity, so it does not change the senior premium, and initial assets and loan flows are transfers relative to the outside benchmark. Borrowing therefore relaxes current liquidity without changing the present-value comparison that governs participation. The floor collects the institutions that keep entry compensation from falling below what a graduate needs to live on: legal limits on sub-minimum and unpaid work, and above all the inability of a graduate without assets to borrow against future human capital, the credit constraint that Becker (1964) identified as the reason workers underinvest in general training and that motivates public student lending. A legal floor and a liquidity floor can bind separately. The floor matters only where it exceeds the price of the machine's output, $\underline{w} > c$ (Proposition 5); for routine cognitive work, where c is close to zero, this is close to automatic. The junior wage is therefore

$$w_t^J \;=\; \max\{\underline{w},\; \omega - \beta m D_{t+1}\}, \tag{6}$$

and the premium is $p_t = w_t^J - c$. At a constant stock, $D_{t+1} = \Delta(S)$; because Δ is decreasing in S, the market-clearing wage $\omega - \beta m \Delta(S)$ is increasing in S: the scarcer are seniors, the more a graduate will pay to become one, and the lower the wage she would accept. Define $\hat{S}$ by

$$\beta m \Delta(\hat{S}) \;=\; \omega - \underline{w}. \tag{7}$$

At a constant stock the floor binds if and only if $S < \hat{S}$: it binds precisely when seniors are scarce enough that the senior premium is large.

Timing and choices. Within a period, firms hire juniors and any residency admits students; production takes place; then juniors mature with probability m and enter the senior market in the next period, newly matured seniors are matched to firms according to Assumption 1(b), and seniors retire with probability δ. Firms choose how many juniors to hire each period and, implicitly, how much routine work to buy from the machine. Graduates choose among a junior position, a residency slot, and the outside option. The senior wage is set by Assumption 1; the junior wage by (6). The intermediary of Section 5, when it exists, chooses the number of slots, the fee, and the tuition, subject to breaking even. The only quantities taken as exogenous are the technology parameters, including the price of the machine, and the wealth and credit of graduates.

2.2 The Pre-AI Economy

Before AI, $c = \infty$: routine work can only be done by juniors. Each firm employs $\bar{J}$ juniors—the requirement leaves no choice—and the pipeline runs itself.

Proposition 1 (Pre-AI economy)**.** *Without the machine, every firm employs $J = \bar{J}$ juniors every period, the steady-state senior stock is $S^{pre} = m\bar{J}/\delta$, and the junior wage is $w^J = \max\{\underline{w}, \omega - \beta m \Delta(S^{pre})\}$. Junior employment, maturation, and the senior stock are independent of the separation probability λ and of the decision-maker's discount factor β^M.*

Three features of this regime deserve emphasis, because AI removes each of them. First, there is no training decision: juniors are hired because the routine work must be done, and maturation follows. Second, the poaching externality is dormant: because the firm would hire $\bar{J}$ juniors regardless of what happens to them afterward, the fraction λ of matured seniors it loses to competitors does not change its behavior. Third, myopia is irrelevant for the same reason. The senior stock S^{pre} is determined by the volume of routine work, not by the value of seniors; this is the pyramid, with many juniors at the base because the base is where the routine work was done.

2.3 The AI Shock and the Firm's Hiring Problem

Generative AI makes c finite. Nothing else changes: juniors are as productive at routine work as before, and the learning technology m is untouched. What changes is that a junior's

routine work now has a market alternative. Firm i chooses $\{J_{i,t}\}$ to maximize

$$\sum_{t\geq 0}(\beta^M)^t\Big[G(S_{i,t}) - w_t^S S_{i,t} - c\bar{J} - p_t J_{i,t}\Big] \tag{8}$$
$$\text{subject to} \quad S_{i,t+1} = (1-\delta)S_{i,t} + (1-\lambda)\,mJ_{i,t} + P_{i,t}, \quad 0 \leq J_{i,t} \leq \bar{J},$$

taking the wage paths and the inflow $P_{i,t}$ of seniors trained elsewhere as given. If $p_t < 0$, juniors remain cheaper than the machine; if $p_t > 0$, each junior costs the firm more than the machine would, and the only reason to employ one is that the junior may become a senior.

Lemma 1 (Hiring rule)**.** *Let $Q_{i,t} \equiv \sum_{s\geq 0}(\beta^M(1-\delta))^s\,[G'(S_{i,t+s}) - w^S_{t+s}]$ be the marginal value of a senior to firm i. The problem (8) is concave, and $J_{i,t}$ is optimal if and only if*

$$p_t - \beta^M(1-\lambda)\,m\,Q_{i,t+1} \begin{cases} \geq 0 & \text{if } J_{i,t} = 0, \\ = 0 & \text{if } 0 < J_{i,t} < \bar{J}, \\ \leq 0 & \text{if } J_{i,t} = \bar{J}. \end{cases}$$

In a symmetric equilibrium $Q_{i,t} = Q_t$ of (3).

Prices at each date depend on next period's senior values, not on the current stock: a junior hired now is priced from the senior market she will enter. Aggregate consistency requires $S_{t+1} = (1-\delta)S_t + mJ_t$ per firm.

Definition 1 (Equilibrium)**.** Given $S_0 = S^{\text{pre}}$, a perfect-foresight equilibrium is a sequence $\{S_t, J_t, w_t^S, w_t^J, D_t, Q_t\}_{t\geq 0}$ with $S_{t+1} = (1-\delta)S_t + mJ_t$, $w_t^S = \theta G'(S_t)$, D_t and Q_t the bounded solutions of (3), graduates' consumption and occupation plans feasible under Assumption 2, w_t^J given by (6), and $J_t \in [0, \bar{J}]$ satisfying the conditions of Lemma 1 with Q_{t+1} in place of $Q_{i,t+1}$. A steady state is an equilibrium in which S_t is constant.

In a steady state $D = \Delta(S)$, $Q = v^M(S)$, $J = \delta S/m$, and the interior hiring condition is a single equation in S,

$$p(S) \;=\; \beta^M(1-\lambda)\,m\,v^M(S), \qquad p(S) = \max\{\underline{w},\, \omega - \beta m \Delta(S)\} - c. \tag{9}$$

The left side is the premium the firm pays for a junior; it is weakly increasing in S, because the more abundant are seniors, the less a graduate will pay for training. The right side is the decision-maker's valuation of the expected retained senior a junior produces; it is strictly decreasing in S and unbounded as $S \to 0$. Let $\Lambda(S)$ denote the right side minus the left

side, and let

$$\hat{c} \;\equiv\; \max\{\underline{w}, \omega - \beta m \Delta(S^{\text{pre}})\} - \beta^M (1-\lambda) m v^M(S^{\text{pre}}).$$

Proposition 2 (The AI shock)**.** *For any $c \in [0, \omega)$ and $0 < \beta^M \leq \beta$, (9) has a unique root $S^e > 0$.*

(i) If $c \geq \hat{c}$, then $S^e \geq S^{pre}$ and the constant path $S_t = S^{pre}$, $J_t = \bar{J}$ is an equilibrium: juniors are not displaced.

(ii) If $c < \hat{c}$, then $S^e < S^{pre}$ and the coast-and-maintain path $S_t = \max\{(1-\delta)^t S^{pre}, S^e\}$ is an equilibrium. No firm hires juniors until the stock reaches S^e; in the period in which it would fall below S^e, firms hire just enough to reach it; thereafter they hire $\delta S^e/m$ per period.

(iii) The steady-state stock is increasing in c: the cheaper the machine, the thinner the pipeline. The senior wage rises above its pre-AI level, and the market-clearing junior wage $\omega - \beta m \Delta(S^e)$ falls below its pre-AI level.

(iv) $S^e > 0$ for every $c < \omega$ and every $\beta^M > 0$.

The threshold $\hat{c}$ makes precise what "AI is cheaper than a junior" means for hiring. It is not $c < w^J$: because a junior may become a senior worth $v^M(S^{\text{pre}})$ to the firm, juniors are displaced only when the machine undercuts the junior wage by more than the capitalized training value. Part (ii) is where time-indexed values matter. Along the declining path above the target, future senior wages and rents are below their eventual stationary levels, so the value of a junior's maturation is below its steady-state level, while the premium is at least its steady-state level; hiring is unattractive until the stock reaches the target, and at the target the stationary conditions apply. Part (iii) is the equilibrium response that a model with a fixed senior surplus cannot deliver. Part (iv) says that the pipeline never vanishes: as seniors become scarce their price rises, and at some stock even a heavily discounted rent covers the premium. The proof characterizes this equilibrium; I do not claim it is the only one.

3. Equilibrium Analysis

The steady state in (9) is inefficient for three reasons that I now separate: firms count only the fraction $1-\lambda$ of matured seniors they retain; when the floor binds, they cannot collect the worker's valuation of training; and their decision-makers may discount at $\beta^M < \beta$. I begin with the planner's benchmark, then remove the floor and myopia to isolate the poaching wedge, and then reintroduce each.

3.1 The Planner's Benchmark

Wages are transfers. The resources at stake are the machine's cost, the routine-work time of juniors, whose social cost is their outside option ω, the time of seniors, valued likewise at ω, and senior output. The planner sets $\lambda = 0$ because separations are transfers among firms, verifies training at no cost, and chooses $\{J_t\}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t N\big[\, G(S_t) - \omega S_t - c\bar{J} - (\omega - c) J_t \,\big] \quad \text{subject to} \quad S_{t+1} = (1-\delta)S_t + mJ_t, \quad J_t \in [0, \bar{J}]. \tag{10}$$

Substituting the law of motion into the objective delivers a useful representation. Define

$$\Phi(S) \;\equiv\; G(S) - \omega S - c\bar{J} - (\omega - c)\,\frac{1 - \beta(1-\delta)}{\beta m}\, S. \tag{11}$$

Proposition 3 (Planner's benchmark)**.** *Up to a constant that depends only on the initial stock, the planner's objective equals $N \sum_t \beta^t \Phi(S_t)$. The function Φ is strictly concave with a unique maximizer S^* characterized by*

$$\omega - c \;=\; \beta m\big[\Delta(S^*) + v(S^*)\big] : \tag{12}$$

the social cost of a junior relative to the machine equals the expected discounted social value of the senior the junior may become. If $S^* < S^{pre}$, the optimal path is the most rapid approach to S^*: no juniors are hired while the stock exceeds S^*, hiring in the crossing period brings the stock to S^*, and $\delta S^*/m$ are hired thereafter. If $S^* \geq S^{pre}$, the planner holds the stock at S^{pre}. AI displaces juniors in the first best—$S^* < S^{pre}$—if and only if $c < c^* \equiv \omega - \beta m[\Delta(S^{pre}) + v(S^{pre})]$, and $c^* \leq \hat{c}$, strictly if $\lambda > 0$ or $\beta^M < \beta$.

The first best after AI is a thinner pipeline than before, provided the machine is cheap enough: this is the diamond that replaces the pyramid. The pre-AI stock was determined by how much routine work there was; the efficient post-AI stock is determined by how valuable seniors are. That $c^* \leq \hat{c}$ says the market displaces juniors at machine prices at which the planner would not, and previews the direction of the inefficiency. The target S^* is the modified golden rule, and the most-rapid-approach property, familiar from Spence and Starrett (1975), makes welfare comparisons across regimes tractable in Section 3.5.

Assumption 3 (Interior comparison region)**.** $c^* > 0$ and $0 \leq c < c^*$.

I maintain Assumption 3 for the rest of the paper. Since $\hat{c} \geq c^*$ in every direct-market regime below (the proof of Proposition 3 applies with the floor and discount factor of that

regime), it implies that juniors are displaced in every regime, so that all stocks are strictly below S^{pre}. Without it the hiring cap must be retained and some rankings become equalities. The assumption holds in the numerical illustration.

3.2 Flexible Wages: The Becker Benchmark

Suppose the floor is slack, so that $w^J = \omega - \beta m \Delta(S)$, and set $\beta^M = \beta$. The steady state condition (9) becomes

$$\omega - c \;=\; \beta m \big[\Delta(S^B) + (1-\lambda)\, v(S^B)\big]. \tag{13}$$

Proposition 4 (Flexible wages)**.** *With flexible wages and patient firms, the unique steady state S^B satisfies $S^B \le S^*$, with equality if and only if $\lambda = 0$; as $\theta \to 1$, $S^B \to S^*$. The shortfall is increasing in λ and decreasing in θ. With flexible wages and a myopic decision-maker, the steady state $S^{B,M}$ solves*

$$\omega - c \;=\; \beta m \Delta(S) + \beta^M (1-\lambda) m v^M(S), \tag{14}$$

is increasing in β^M, and is bounded below, for every $\beta^M \ge 0$, by the stock $\tilde{S}$ that workers alone would finance, $\beta m \Delta(\tilde{S}) = \omega - c$; as $\beta^M \downarrow 0$, $S^{B,M} \to \tilde{S}$. With isoelastic production, $G'(S) = AS^{-\eta}$, $\tilde{S}/S^* = \theta^{1/\eta}$.

Comparing (13) with (12), the only wedge is $(1-\lambda)$ applied to the firm's share of the return. Workers pay for their own share in full through the wage: the worker's valuation $\Delta(S)$ enters the market condition exactly as it enters the planner's. In the limit in which firms earn no rent on seniors ($\theta \to 1$), the entire return accrues to the worker, the worker finances the entire training, and the outcome is efficient for any λ—Becker's (1964) result. Poaching matters only to the extent that wage compression gives firms a stake in general human capital, as in Acemoglu and Pischke (1999), and the loss it causes is confined to the firm's share.

The second part of the proposition is the reason myopia, on its own, is limited in its damage. If the decision-maker undervalues the future, the junior wage falls to compensate; workers, who value the future at β, absorb what the firm will not pay. Even a decision-maker with no horizon at all ($\beta^M = 0$) would hire juniors up to the stock at which workers' willingness to pay covers the full premium. Limited is not zero: worker financing alone delivers the fraction $\theta^{1/\eta}$ of the efficient stock, which is 0.68 at the parameters of the illustration in Section 5.6, and lower where firms retain more of the senior's product. There is no extinction

with flexible wages, because the agents who bear the consequences of a thin pipeline are the ones who buy the training.

3.3 The Liquidity Floor

The market-clearing junior wage $\omega - \beta m D_{t+1}$ is increasing in the senior stock, and AI reduces the senior stock. As the pipeline thins, the senior premium rises, graduates are willing to accept lower entry wages, and the wage they would accept eventually falls through the floor $\underline{w} = \ell - a - b_F$, which the private market will not let them borrow beneath. At the floor-constrained steady state of the numerical illustration in Section 5.6 that hypothetical clearing wage is negative, but it is not the wage of a flexible economy. If graduates can fund current consumption, the stock is larger, the senior premium smaller, and the market clears at a positive wage: 0.41 in the illustration, against a floor of 0.60. The binding problem is the inability to smooth consumption while accepting a lower positive wage, not a necessary negative wage in the flexible equilibrium. Proposition 8 quantifies the credit that closes the gap.

Proposition 5 (Liquidity floor)**.** *Let $\hat{S}$ be defined by (7), so that the floor binds at a constant stock S if and only if $S < \hat{S}$.*

(i) The floor binds in the post-AI steady state if and only if $\underline{w} - c > \beta^M(1-\lambda)mv^M(\hat{S})$. In particular, it cannot bind if $\underline{w} \le c$. It is slack before AI whenever $S^{pre} > \hat{S}$.

(ii) When the floor binds in the flexible-wage benchmark, $S^B < \hat{S}$, the steady state S^F (for $\beta^M = \beta$) solves

$$\underline{w} - c \;=\; \beta(1-\lambda)\, m\, v(S^F), \tag{15}$$

is strictly below S^B, is decreasing in $\underline{w}$, and is independent of the worker's valuation Δ.

(iii) Consequently, when the floor strictly binds, the steady-state stock is increasing in graduates' assets a and private credit b_F and decreasing in living costs ℓ, wherever $\underline{w} = \ell - a - b_F > w_{\min}$.

The floor severs the link between the worker's valuation of training and the firm's decision. In (15) the worker's side of the return has disappeared: the firm pays $\underline{w}$ regardless of what the graduate would accept, and hires only if its own discounted rent covers the premium $\underline{w} - c$. The premium is now a parameter, the gap between the floor and the price of the machine, and the two frictions that the flexible wage neutralized are exposed. The condition in part (i) is naturally met after AI and not before. Before AI the stock is large, the senior premium modest, and the markdown graduates accept small; after AI the stock is small

and the markdown they would accept large. The same institution that was innocuous when routine work was scarce becomes binding when routine work is cheap. Part (iii) is the empirical content of the liquidity interpretation: junior hiring in AI-exposed occupations should fall more where entry-level workers have less wealth and less access to credit, and less where parents or lenders can carry them.

3.4 Managerial Myopia

The decision-maker discounts at $\beta^M < \beta$. This is a reduced form. Managerial short-termism has standard microfoundations in corporate finance—executives who expect to be replaced, and compensation tied to near-term earnings, weight the near term more heavily than a diversified shareholder would (Stein, 1989)—and the survey evidence that a majority of executives would sacrifice long-run value to smooth reported earnings (Graham, Harvey, and Rajgopal, 2005) suggests that investments with a long payoff, of which developmental hiring is a leading example, are exactly the ones cut. The reduced form requires the additional assumption that the decision-maker's payoff, and not merely shareholders' value, drives the hiring decision. With flexible wages myopia lowers the stock but workers pick up much of what the firm will not pay (Proposition 4); the floor removes that worker-financing feedback and so amplifies myopia.

Proposition 6 (Myopia under the floor)**.** *When the floor binds, the steady state S^M solves*

$$\underline{w} - c \;=\; \beta^M (1-\lambda)\, m\, v^M(S^M), \tag{16}$$

is strictly below S^F if $\beta^M < \beta$, and is increasing in β^M. If $\underline{w} > c$, the floor binds for sufficiently small β^M and $S^M \to 0$ as $\beta^M \downarrow 0$; for every $\beta^M > 0$, $S^M > 0$. With isoelastic senior production, $G'(S) = AS^{-\eta}$, the two frictions compound as

$$\frac{S^M}{S_0^F} \;=\; \left[(1-\lambda)\,\frac{\beta^M\,[1-\beta(1-\delta)]}{\beta\,[1-\beta^M(1-\delta)]}\right]^{1/\eta},$$

where S_0^F is the floor-constrained stock absent poaching and myopia, $\underline{w} - c = \beta m v(S_0^F)$.

Two remarks are in order. First, extinction requires both frictions. The pipeline goes to zero only in the limit of a decision-maker who counts nothing beyond the current period, and only when the floor prevents workers from stepping in; for any $\beta^M > 0$ the rising senior premium ensures a positive pipeline. The result is therefore not that firms collectively allow the supply of seniors to disappear, but that they run a pipeline that is too thin, by a factor that compounds the poaching share, the discount on the period of maturation,

and the shortened horizon over which the senior's rent is capitalized. Second, the myopic decision-maker's behavior is individually rational at every step: each firm is indifferent at the margin between a junior and the machine, and the collective outcome is the sum of these margins. Governance mechanisms that lengthen horizons—long-vesting compensation, patient ownership—enter the model through β^M and are one of its testable margins.

3.5 Welfare, With and Without AI

Every direct-market regime studied so far, and the intermediary of Section 5 with a constant supervision cost and ample capacity, produces the same shape of path from the pre-AI stock: the senior stock coasts downward with no developmental hiring until it reaches the regime's steady state, and is then maintained (Proposition 2). For $s \in [0, S^{\text{pre}}]$ define the coast-and-maintain path $S_t(s) = \max\{(1-\delta)^t S^{\text{pre}}, s\}$ with $J_t(s) = [S_{t+1}(s) - (1-\delta)S_t(s)]/m$, and let $W(s)$ be the value of (10) along it, net of any supervision cost where stated. This notation denotes a specified allocation; Proposition 2 establishes when such an allocation is a direct-market equilibrium. Write $W^* = W(S^*)$, $W^B = W(S^B)$, $W^F = W(S^F)$, and $W^M = W(S^M)$. The representation in Proposition 3 makes these paths directly comparable.

Proposition 7 (Welfare ranking). *(i) For $0 \le s_a < s_b \le S^*$, $W(s_a) < W(s_b)$. Accordingly*

$$W^M \le W^F \le W^B \le W^*.$$

The first inequality is strict if $\beta^M < \beta$ and the floor binds; the second is strict if the floor strictly binds for patient firms; the last is strict if $\lambda > 0$.

(ii) AI raises first-best welfare, $W^* > W^{pre}$, if and only if $c < c^*$.

(iii) If $W(0) < W^{pre}$, there is a unique $\underline{S} \in (0, S^*)$ such that $W(s) < W^{pre}$ if and only if $s < \underline{S}$. The decentralized response to AI can leave the economy worse off than never adopting AI, even though the first-best response adopts it.

Part (iii) is the sense in which the ladder is broken. The technology cannot lower first-best welfare, because the first best can retain the old allocation; a planner who adopts the machine and thins the pipeline to S^* is better off than before. A market that thins the pipeline to S^M can be worse off, because the routine-work saving is a flow while the senior stock is a capital good, and the market runs the capital good down. The loss arises from the decentralized financing response to an expanded production set, and its occurrence and size depend on the financing, appropriation, and technology parameters. The comparison is not between AI and no AI but between two responses to AI, one of which finances the pipeline and one of which does not.

4. Credit and Transfers Aimed at Firms

There are several ways to finance the pipeline, and an intermediary is one possible response to a restricted contract menu; it is not a substitute for establishing what ordinary credit and transfer instruments can accomplish. This section studies those instruments. All arrangements are evaluated with the same accounting: real resources (machine output, workers' time, and any verification or supervision) count in welfare, while wages, loan repayments, and transfers cancel.

4.1 Credit for Conventional Apprenticeship

When the liquidity term in $\underline{w}$ binds, raising b_F lets workers accept lower current wages, and it does so without any resource cost. Let $w^{B,M} \equiv \omega - \beta m \Delta(S^{B,M})$ be the market-clearing junior wage at the flexible-compensation myopic stock of (14).

Proposition 8 (Credit reform)**.** *If $w_{\min} \leq w^{B,M}$ and $b_F \geq \ell - a - w^{B,M}$, with (5) holding for b_F, then $\underline{w} \leq w^{B,M}$ and the direct market attains the flexible-compensation stock $S^{B,M}$, with no subsidy, no intermediary, and no resource cost. The analogous statement holds for patient firms at S^B.*

Because the floor is slack at $S^{B,M}$, the hiring condition (9) has the same root as in the flexible benchmark. In the illustration of Section 5.6 the market-clearing wage at $S^{B,M}$ is 0.41, so raising conventional credit from $b_F = 0.10$ to 0.29 (to 0.22 for patient firms) lifts the stock from 1.34 to 3.16 and welfare from -22.45 to $+3.87$ relative to the pre-AI economy. The reform leaves the turnover wedge on the employer's share and any managerial discounting of it, so it does not reach the first best. Because b_F is a primitive that reflects enforceability, the reform is an institutional change—for example, government-guaranteed income-contingent credit that extends to trainees at accredited employers—rather than a parameter the government sets.

4.2 Transfers to Firms

For a transparent implementation benchmark, suppose the government can verify developmental apprenticeship at no additional resource cost, and that all transfers below are financed by lump-sum taxes on an unconstrained outside sector, which do not reduce trainee assets or borrowing limits. Section 4.3 relaxes both assumptions.

Proposition 9 (Direct implementation)**.** *Under Assumption 3 and the stated verification and financing benchmark, steady-state implementation of S^* is as follows.*

(i) A payment of $v^M(S)$ to the training firm for each newly qualified worker assigned elsewhere, financed by the receiving firm, removes the retention factor $1-\lambda$ from the marginal training return. With flexible compensation and $\beta^M = \beta$ it implements S^*. It does not remove a binding compensation floor or the managerial discount wedge. Matching and the wage rule are held fixed; the proposition does not optimize labor mobility.

(ii) A subsidy per verified apprenticeship, paid in the training period, implements S^* at

$$s^* \;=\; p(S^*) - \beta^M(1-\lambda)\, m\, v^M(S^*) \;\geq\; 0, \tag{17}$$

which is finite for every $\beta^M \geq 0$.

(iii) A payment per successful qualification, made at the beginning of the following period, implements the same target at

$$r^* \;=\; \frac{p(S^*)}{\beta^M m} - (1-\lambda)\, v^M(S^*) \;=\; \frac{s^*}{\beta^M m}. \tag{18}$$

It is decreasing in β^M and, for the maintained $\theta < 1$, unbounded as $\beta^M \downarrow 0$.

The government's date-t expenditure is Ns^*J_t for the current subsidy and Nr^*mJ_{t-1} for the qualification payment, with no inherited payment obligation at the initial date unless explicitly assigned. Both budgets must be funded. Part (i) says that the Pigouvian remedy for poaching is available and, absent the floor, sufficient; a receiving firm taxed exactly its marginal value of a new senior is indifferent to accepting that marginal match, and the matching protocol selects acceptance, so the transfer is an accounting implementation under fixed matching, not a result about endogenous recruitment. Parts (ii) and (iii) show that myopia makes delayed incentives expensive but does not make policy toward firms impossible: the current subsidy is effective whatever the manager's horizon, and the qualification payment is the same incentive delivered late, so it must scale with $1/\beta^M$. Nor does $\beta^M \to 0$ alone prove that a firm rejects a scheme whose promised payment grows as $1/\beta^M$: the product $\beta^M r^*$ need not vanish. A voluntary arrangement requires a separate specification of levies, timing, access to the common pool, and nonparticipant payoffs, as in the participation questions studied by Jackson and Palfrey (2001). The formulas concern mandatory, externally financed implementation.

4.3 Verification and a Common Comparison

An employment record may be easier to verify than the developmental content of a position. This observation limits what a subsidy conditional only on head count can guarantee, without establishing that all such subsidies leave training unchanged. If a firm can assign a hired junior to work that generates no qualification, and workers cannot tell developmental from non-developmental work when they accept a wage, the subsidy can be collected without buying training, and anticipated diversion changes the wage and training equilibrium in ways this model does not solve. If workers do observe assignment, diversion is constrained by their participation: a non-developmental employee, who receives no future training benefit, must be paid at least ω, so one cannot compare developmental and non-developmental jobs while holding the training-compensating wage fixed. I therefore claim only that an unconditional hiring subsidy is exposed to diversion, not that it is ineffective.

The analysis below therefore introduces an explicit certification technology for a specified human-practice curriculum. Certification records trainees' performance of the required tasks and lets payment and financing eligibility be conditional on that performance. Verifying only the final deliverable would not suffice, because the machine can produce it without training a human. Monitoring is provided under an enforceable service contract, and its real resource cost is included in the program cost. This is a maintained monitoring technology, not a result that university ownership makes training observable. An accredited employer or professional association may purchase the same monitoring. If direct employer verification costs κ_F per trainee, that cost must enter both the firm's premium and the social benchmark, and the subsidy should not be read as eliminating it. Likewise, an accredited employer program that obtains the same worker-financing terms is an alternative to a stand-alone intermediary. The core residency comparison therefore holds the contract menu fixed: conventional employer apprenticeships carry the borrowing limit b_F, and the specified accredited educational contract carries b_U. Endogenous entry into integrated employer–education contracts is an additional margin, not ruled out by a claim that only universities can organize training.

5. The Training Intermediary

The agents who value a senior career at β are the graduates. The market cannot collect their valuation through the wage because a junior's compensation cannot fall below $\underline{w}$, and credit reform may not be available. An intermediary that stands between graduates and firms can collect the valuation as tuition, provided the price is not a wage, students can finance it, and the training it sells is credible. I model such an intermediary as a university

residency, modeled on the medical residency and clinical legal education and generalized to knowledge-work professions, and I ask what it can and cannot achieve.

5.1 The Mechanism

Definition 2 (Residency)**.** A stand-alone accredited provider (a) enrolls $J_t^R \leq \bar{J}^R$ residents per firm each period, where $\bar{J}^R \leq \bar{J}$ is the capacity set by available projects and certified supervisors and $J_t + J_t^R \leq \bar{J}$; (b) assigns each resident to personally complete one junior's worth of a sponsoring firm's routine work under supervision, for which the firm pays a fee f_t; (c) charges each resident a net tuition τ_t, tuition less any stipend; and (d) incurs a supervision, certification, and project-sourcing cost $\kappa \geq 0$ per resident and receives a public subsidy $g \geq 0$ per verified resident, financed by nondistortionary taxes. A resident qualifies with probability m. Qualified residents enter the common senior matching pool, with no privileged assignment to their sponsoring firms or to the provider.

Assumption 4 (Intermediary)**.** (a) *Financing.* The educational arrangement meets the requirements for its compensation and tuition treatment and qualifies for the borrowing limit b_U, so net tuition is feasible only if $\tau_t \leq a + b_U - \ell$. (b) *Credible training.* The certification technology of Section 4.3 commits the provider to actual human practice; substituting an AI-produced deliverable cannot satisfy it. The provider can commit to maturation probability m through accreditation, supervised process records, and assessment, whose cost is in κ. (c) *Break-even.* The provider is competitive or operates at break-even prices and cannot divert earmarked payments while retaining certification.

The three parts are conditions a program must meet, not automatic consequences of a university's involvement. Part (a) is an institutional fact about educational status and credit. Whether a resident is a student or an employee of the sponsoring firm turns on facts about the program: the Department of Labor's guidance applies a fact-specific primary-beneficiary test to interns and students at for-profit employers, rather than affiliation alone (U.S. Department of Labor, 2018), and federal student aid is available for enrollment in an eligible degree or certificate program (Federal Student Aid, n.d.), which a newly proposed micro-residency does not automatically satisfy. The model treats b_U and the legal status of net tuition as properties of a qualifying program; it does not establish that any proposed residency is legally or financially eligible. Part (b) is the counterpart of the observation in Section 4.3 that a hiring subsidy cannot verify assignment. A school does not solve the problem merely by controlling tasks, because a firm also controls the tasks it assigns, and a deliverable does not reveal whether a resident learned: the machine could supply the

same output. Credibility must come from a commitment technology, and accreditation and supervised process records are candidates whose cost is κ. If credibility rested instead on the institution's reputation, the institution would need a horizon, and I would not be entitled to say that it needs none. Part (c) says the intermediary responds to demand: it has no welfare objective and no patience beyond that of the graduates it serves.

Because residents and conventional apprentices use the same finite routine requirement, $J_{i,t} + J^R_{i,t} \leq \bar{J}$; and because qualified residents enter the common pool, senior inflows are $P_t = \lambda m J_t + m J^R_t$ and the aggregate law is

$$S_{t+1} \;=\; (1-\delta)S_t + m\big(J_t + J^R_t\big). \tag{19}$$

The constant cost κ applies within the capacity bound $\bar{J}^R$; it does not assert an unlimited supply of senior supervisors or live projects. Given residency allocations, the firm's problem has residual capacity $\bar{J} - J^R_{i,t}$ and an additional flow term $-(f_t - c)J^R_{i,t}$, which is zero at $f_t = c$; one firm's sponsorship has no effect on its own future matching inflow in the continuum economy.

The sponsoring firm receives a deliverable worth c, the price of the machine output it replaces, and participates only if $f_t \leq c$. A graduate who enters a residency forgoes ω for one period, pays τ_t, and with probability m enters a senior career worth D_{t+1} in excess of the outside option, so she participates if and only if $\tau_t \leq \bar{\tau}_t \equiv \beta m D_{t+1} - \omega$; her current financing constraint is $\tau_t \leq \tau^L \equiv a + b_U - \ell$, because she must fund both tuition and living expenses. A slot breaks even if and only if $f_t + \tau_t + g \geq \kappa$. What the floor prevents the firm from collecting through the wage, the intermediary collects through tuition, subject to liquidity. Taking $f_t = c$, a break-even provider charges $\tau^0 = \kappa - c - g$, and a slot is feasible only if both

$$\kappa - c - g \;\leq\; a + b_U - \ell, \tag{20}$$

$$\beta m D_{t+1} \;\geq\; \omega - c + \kappa - g. \tag{21}$$

Willingness to pay and ability to pay are separate: a large expected senior premium does not relax (20). Define S^R by

$$\beta m \Delta(S^R) \;=\; \omega - c + \kappa - g, \tag{22}$$

and assume $\omega - c + \kappa - g > \beta m \Delta(S^{\text{pre}})$ so that $S^R \in (0, S^{\text{pre}})$ is interior; if the inequality is reversed, all stocks up to S^{pre} meet the valuation condition and capacity, not an unrestricted root, limits enrollment. Because Δ is decreasing in S, the intermediary can finance slots

only when seniors are scarce enough that the senior premium supports the tuition, which is the same scarcity that drives the market wage through the floor. A residency equilibrium augments Definition 1 with these contracts, budgets, capacity constraints, and provider supply conditions. Providers fill financially feasible, privately acceptable slots up to capacity, and excess demand for such slots is rationed.

5.2 Equilibrium with a Residency

Two welfare benchmarks are needed. The planner of Section 3.1 verifies training at no cost and reaches S^*. If training can only be delivered through the residency, the efficient stock $S^{*,R}$ (assuming capacity is sufficient) solves

$$\omega - c + \kappa \;=\; \beta m\big[\Delta(S^{*,R}) + v(S^{*,R})\big], \tag{23}$$

which is below S^* because supervision is a real cost. Let $W^{*,R}$ denote welfare net of supervision at $S^{*,R}$. For an active residency on a coast-and-maintain path define $L(s) \equiv \sum_{t\geq 0} \beta^t J_t(s)$ and

$$W^R(\kappa, g) \;=\; W(S^R) - N\kappa L(S^R). \tag{24}$$

Subsidies and tuition are transfers; supervision, certification, and project-sourcing resources are real costs and remain in welfare regardless of who pays them.

Proposition 10 (Residency)**.** *Suppose S^R is defined by (22) and interior, let S^M denote the steady state of (9) (which solves (16) when the floor binds), and let $\bar{J}^R$ denote capacity.*

(i) The firm's participation constraint $f_t \leq c$ is static and involves neither β^M nor other firms' actions. A resident's constraints depend on her anticipated senior wage path through D_{t+1} and on liquidity (20), but not on β^M.

(ii) A privately acceptable residency expands the stationary stock only if (20) holds and $S^R > S^M$, which is equivalent to

$$\omega - \beta m \Delta(S^M) \;<\; c + g - \kappa : \tag{25}$$

the wage graduates would accept at the market steady state is below the price of the machine plus the net subsidy, net of supervision. If $S^R < S^M$ it is inactive; at equality allocations can mix, and I select the inactive residency when referring to a single welfare value.

(iii) (Ample capacity.) If $S^R > S^M$, (20) holds, and $\bar{J}^R \geq \delta S^R/m$, the stock is S^R, conventional developmental hiring is zero, $J^R = \delta S^R/m \leq \bar{J}$, and residents pay net tuition

$\tau^0 = \kappa - c - g$. *The coast-and-maintain path at* S^R *is a perfect-foresight equilibrium.*

(iv) (Limited capacity.) If $S^R > S^M$ *and* (20) *holds, the stationary stock and flows are*

$$S = \max\Big\{S^M,\ \min\big\{S^R,\ m\bar{J}^R/\delta\big\}\Big\}, \tag{26}$$

$$J^R = \min\{\bar{J}^R,\ \delta S/m\}, \qquad J = \delta S/m - J^R. \tag{27}$$

In particular, capacity below $\delta S^M/m$ *displaces conventional training one for one without raising the stock. These are stationary results; no coast-and-maintain transition is asserted when capacity binds.*

(v) $S^{*,R} < S^*$ *if* $\kappa > 0$. *Where the root of* (22) *is interior,* $S^R < S^*$ *if and only if* $g < \kappa + \beta m v(S^*)$; $S^R < S^{*,R}$ *if and only if* $g < g^{*,R} \equiv \beta m v(S^{*,R})$; *and* $S^R < S^B$ *if and only if* $g < \kappa + \beta(1-\lambda) m v(S^B)$. *With no subsidy the residency stops short of the first best, of the residency-efficient stock, and of the flexible-wage market.*

(vi) The subsidy $\kappa + \beta m v(S^*)$ *implements the stock* S^* *but not first-best welfare: welfare on that path is* $W^* - N\kappa L(S^*) < W^*$. *The subsidy* $g^{*,R}$ *implements* $S^{*,R}$ *and maximizes welfare given the residency technology, provided credit* (20) *and capacity do not bind. If conventional apprenticeships remain available, implementation through an active residency also requires* $S^{*,R} > S^M$, *so that direct training is displaced; otherwise the residency-only optimum is a conditional benchmark, not the outcome of the mixed market.*

(vii) Let $g = 0$ *and* $\kappa^{act} \equiv \beta m \Delta(S^M) - \omega + c > 0$. *Suppose capacity is sufficient at* $S^R(0)$ *and* $a + b_U - \ell + c \geq \kappa^{act}$, *so that student financing is feasible for all* $0 \leq \kappa < \kappa^{act}$. *On this interval* W^R *is continuous and strictly decreasing in* κ, *and* $W^R > W^M$ *at* $\kappa = 0$. *There is a unique* $\bar{\kappa} \in (0, \kappa^{act})$ *with* $W^R(\bar{\kappa}) = W^M$, *so the residency improves on the market if and only if* $\kappa < \bar{\kappa}$ *on the active branch. As* $\kappa \uparrow \kappa^{act}$, *active-program welfare tends to* $W^M - N\kappa^{act} L(S^M) < W^M$; *once the program is inactive, welfare is* W^M, *and no continuity across that change in organizational form is claimed. For* $\kappa \in [\bar{\kappa}, \kappa^{act})$ *the residency is privately feasible and active yet welfare-reducing.*

Part (i) is the sense in which the mechanism is robust to myopia, and it should be read for what it is: a firm that receives a deliverable worth at least the fee accepts, whatever its horizon, and a graduate who is offered training at a price below its value to her accepts, because she values the future at β. The content of the result lies in the financing the constraints permit. Part (ii) gives the conditions under which the intermediary's financing channel has room to operate. It requires that the senior premium at the market steady state be large enough that graduates would pay at least $\kappa - c - g$ for the training, and that student credit cover living costs and net tuition. Substituting (16) for $G'(S^M)$ when the floor binds,

the first condition reads

$$(\underline{w} - c)\,\frac{\theta\,\beta\,[1-\beta^{M}(1-\delta)]}{(1-\theta)\,\beta^{M}(1-\lambda)\,[1-\beta(1-\delta)]} > \omega\Big(1+\frac{\beta m}{1-\beta(1-\delta)}\Big) - c + \kappa - g,$$

which is more easily satisfied the higher the floor relative to the machine, the larger the workers' share θ, and the shorter the firms' horizon: each thins the market pipeline and raises the senior premium that funds the tuition. It fails when the floor is low, which is when graduates' wealth and credit are ample.

Parts (iii) and (iv) are crowding out and scale. When capacity is ample the crowding out is complete: the intermediary maintains a stock at which the firm's discounted rent no longer covers the premium, so firms stop developmental hiring and the residency becomes the profession's pipeline. When capacity is limited, the residency displaces firm hiring one for one until capacity exceeds $\delta S^{M}/m$, and only capacity beyond the market's replacement flow raises the stock; a smaller program attracts students and wastes κ per resident without adding expertise.[8] The medical residency is the route to practice rather than a supplement to it.

Part (v) locates the mechanism in the welfare ranking. The intermediary cannot collect the rent $v(S)$ that accrues to whichever firm employs the seniors it produces; from its perspective the poaching rate is one. Unsubsidized, it therefore undershoots the first best by that rent and undershoots the flexible-wage market, in which firms collect the fraction $1-\lambda$ of it. Part (vi) separates two things that are easily conflated. A subsidy is a transfer and does not erase the resources spent on supervision, so the subsidy that implements the stock S^{*} does not restore first-best welfare, and it overcorrects relative to what the residency technology can efficiently deliver: with $\kappa = 0.25$ in the illustration below, $S^{*} = 3.76$ but $S^{*,R} = 3.05$, and the subsidies that implement them are 0.61 and 0.43. Part (vii) is a caution about the intermediary's existence. Because entry is privately profitable but crowds out firm hiring, a residency can be active and yet reduce welfare, when its supervision cost is high enough to waste more than the increment in the stock is worth. Together the parts deliver, when the floor binds, firms are myopic, there is no subsidy, capacity is ample, and $\kappa < \bar{\kappa}$, the ranking

$$S^{M} < S^{R} < S^{B} \leq S^{*}, \qquad W^{M} < W^{R} < W^{B} \leq W^{*}, \tag{28}$$

[8] Coexistence of firms and residents at a stock above S^{M} would require convex firm-level costs of absorbing juniors, which I do not model; with the linear hiring technology firms target a stock, and above S^{M} they strictly prefer not to hire. Coexistence at S^{M} arises here only because capacity limits the residency.

with the last inequality strict if $\lambda > 0$, while the position of the floor-constrained patient market S^F relative to S^R depends on parameters. The residency is a second-best device for economies in which the floor has disabled worker financing; where wages are flexible it is dominated by the market, which already collects the worker's valuation and part of the firm's.

5.3 Scale-Dependent Supervision

A constant supervision cost within capacity does not model the supply of senior supervisors or of live projects. Suppose instead that the per-resident cost rises with scale, $\kappa(J^R) = \kappa_0 + \kappa_1 J^R$, because supervision draws on seniors whose time is the scarce input and because each additional project is harder to source; business development is performed by senior practitioners whose time is expensive.

Proposition 11 (Scale-dependent supervision)**.** *With $\kappa(J^R) = \kappa_0 + \kappa_1 J^R$ and ample capacity, the steady-state stock of the active residency is the unique root $S^R(\kappa_1)$ of*

$$\beta m \Delta(S) \;=\; \omega - c + \kappa_0 - g + \kappa_1 \, \delta S/m,$$

which is strictly decreasing in κ_1 and κ_0. The residency is active if and only if $S^R(\kappa_1) > S^M$ and student credit covers living costs and net tuition at that scale, $a + b_U \geq \ell + \kappa(\delta S^R/m) - c - g$.

Scale therefore matters in the way the referees suspected: the same program that operates at a stock S^R for small κ_1 operates at a lower stock as κ_1 rises, and if the market stock is not far below it the advantage disappears. With $\kappa_1 > 0$ the transition is no longer coast-and-maintain, because the price of a slot depends on the volume, so I state this result for steady states only.

5.4 Public Funds

Under the maintained benchmark that transfers are financed by nondistortionary taxes, a verified direct subsidy can achieve the same allocation as the residency at the same resource cost, so the residency's role would be undetermined. Its distinctive feature is that an unsubsidized program raises the stock with no public outlay, because students' tuition and the sponsors' fees cover its cost. Suppose each unit of public outlay costs $1 + \phi$ units of resources, $\phi > 0$, while private transfers and student loans satisfying (5) carry no fiscal cost.

Let $\kappa_F = \kappa$, so that direct verification and the residency use the same real resources per trainee, with verification borne by the firm and entering its premium.

Proposition 12 (Public funds)**.** *For any target stock $S \in (S^M, S^R]$ with $g = 0$, attained by the residency through capacity $\bar{J}^R = \delta S/m$ (or by S^R itself under ample capacity), the residency implements S with no public outlay. Implementing S through a verified direct subsidy requires a per-trainee outlay $s(S) = \kappa_F + p(S) - \beta^M(1-\lambda)mv^M(S) > 0$, and so a fiscal distortion $\phi\, s(S)\, \delta S/m$ per firm per period that the residency avoids. The credit reform of Proposition 8 requires neither outlay nor verification.*

With $\phi > 0$ the residency therefore dominates the verified direct subsidy at the same real cost on $(S^M, S^R]$, and where credit reform is feasible it dominates both. In the illustration, implementing the unsubsidized residency stock $S^R = 2.08$ through a direct subsidy would require an outlay of 0.11 per trainee beyond verification cost, about 0.23 per firm per period. The comparison reverses as $\phi \to 0$, and for targets above S^R both routes require subsidies whose relative size depends on parameters.

5.5 Why Universities, and Why Not Only Universities

The comparative advantage in Proposition 10 rests on Assumption 4, and it is worth stating what it does and does not require. It does not require that universities be patient. The intermediary in the model breaks even period by period; the horizon in the mechanism is the graduate's, and the graduate finances training because she, and not the firm's manager, will collect the senior premium. Nor does it require that universities be insulated from the demands of firms. A firm that hires a residency graduate obtains a trained senior at no cost, exactly as it obtains a poached one, and values her at Q_t; the residency satisfies firms' short-horizon demand for deliverables today and their demand for seniors tomorrow, and graduates' long-horizon demand for training. There is no tension for competition among intermediaries to propagate.

What the mechanism does require is an institution that can obtain student credit for the graduate, certify training credibly, and supply supervised projects at sufficient scale, and the university is not the only institution that can. Professional bodies administer supervised-practice requirements in accounting and law; teaching hospitals, which are firms, run medical residencies under accreditation by a licensing body and with public subsidy; employer–university partnerships and specialized training providers charge tuition. An accredited employer that met Assumption 4(a)–(b) would be an intermediary in the sense of the model. The university's leading position rests on features the model makes explicit but does not

guarantee: its charge is by design tuition rather than a wage, public student credit is in most systems tied to accredited educational institutions, certification is its traditional function (Spence, 1973), and it is the existing channel for public subsidy of human capital. A qualified employer program that supplies the same services at lower resource cost may dominate stand-alone intermediation, and expanding conventional apprenticeship credit can dominate creating an additional institution when it relaxes the same binding liquidity constraint more cheaply. The model neither gives universities a screening advantage nor establishes that they have uniquely low costs.

The model is also explicit about the limits. The supervision cost includes the cost of sourcing projects from firms, capacity is finite, and Section 5.3 shows how a rising cost erodes the advantage. Simulated or retrospective projects lower the deliverable's value below c, which raises the net cost the intermediary must recover, and a fidelity gap between supervised practice and consequential decision-making lowers the maturation probability of residents below m, which lowers the tuition the intermediary can charge; a residency with qualification probability m^R replaces m by m^R in its valuation and production conditions. Each of these is a reason the mechanism may need the subsidy of Proposition 10(v)–(vi), not a reason it cannot operate. The relevant benchmark is medicine, where the residency is the profession's pipeline, is run through firms under an intermediary's accreditation, and is publicly subsidized.

5.6 Numerical Illustration

Table 1 illustrates the regimes with isoelastic senior production, $G(S) = AS^{1-\eta}/(1-\eta)$, and the parameters listed in the table note; derivations are in Appendix B. A period is the length of a junior spell and parameters are per period. These are illustrative parameters, not estimates of apprenticeship duration or evidence that the financing restriction binds in any particular profession. Ten percent of junior spells convert into senior positions; seniors retire with probability 0.10; the firm retains a quarter of a senior's marginal product; nearly a third of newly qualified seniors are lost to competitors; the firm's decision-maker discounts at 0.90 against a social discount factor of 0.95; the machine does a junior's routine work for a fifth of the graduate's outside option; graduates need 0.8 per period to live, hold 0.1 in assets, and can borrow 0.1 against a conventional apprenticeship, so the floor is 0.6; and an eligible residency permits borrowing up to 1.5, which satisfies (5) because annual debt service on that loan, 0.079, is below $\omega - \ell = 0.2$. Before AI, each firm employs five juniors and, in steady state, five seniors. Welfare is reported as the present value from the pre-AI stock, in deviation from the pre-AI value, so that the table reads as a comparison of responses to AI.

Table 1: Equilibrium regimes: a numerical illustration.

Regime	Senior stock S	Trainees J or J^R	Junior pay w^J or $-\tau$	Senior wage w^S	Welfare vs. pre-AI $W^r - W^{\text{pre}}$
Pre-AI (no machine)	5.00	5.00	0.77	1.35	0.00
First best	3.76	3.76	—	1.67	+5.31
Flexible wages, patient firms	3.39	3.39	0.48	1.80	+4.78
Flexible wages, myopic firms	3.16	3.16	0.41	1.90	+3.87
Floor, patient firms	2.06	2.06	0.60†	2.62	−7.23
Floor, myopic firms	1.34	1.34	0.60†	3.62	−22.45
Residency, $g = 0$	2.08	2.08	−0.05	2.60	−13.76
Residency, $g = \kappa$	2.56	2.56	0.20	2.22	−9.88
Residency, $g = 0.61$ (implements S^*)	3.76	3.76	0.56	1.67	−11.08
Residency, $g = 0.43$ (implements $S^{*,R}$)	3.05	3.05	0.38	1.95	−8.69

Note: Parameters are $\beta = 0.95$, $\beta^M = 0.90$, $\delta = 0.10$, $m = 0.10$, $\lambda = 0.30$, $\theta = 0.75$, $\omega = 1$, $c = 0.20$, $\ell = 0.8$, $a = 0.1$, $b_F = 0.1$, $w_{\min} = 0$, $b_U = 1.5$ (so $\underline{w} = 0.6$), $\bar{J} = \bar{J}^R = 5$, $\kappa = 0.25$, $A = 6$, $\eta = 0.75$. Senior stock and hiring are per firm. "Junior pay" is the junior wage, or for the residency the stipend net of tuition ($-\tau$). †The floor binds; the wage that would clear the market is −0.06 for patient firms and −0.72 for myopic firms at those stocks. The flexible-wage rows are attained only in an economy with $\underline{w} \leq 0.41$. Welfare is the present value of (10) per firm along the coast-and-maintain path from the pre-AI stock, net of supervision costs for the residency rows, in deviation from the pre-AI present value of 517.8. The supervision cost above which the unsubsidized residency reduces welfare is $\bar{\kappa} = 0.47$, and the cost above which it is inactive is $\kappa^{\text{act}} = 0.92$.

The table reproduces the propositions. The first-best response to the machine thins the pipeline from five seniors to about three and three-quarters and raises welfare by 5.3, about one percent of the pre-AI present value. With flexible wages, the market comes within a tenth of the first-best gain: the poaching wedge costs 0.5, and myopia costs a further 0.9, because the junior wage falls from 0.48 to 0.41 and workers absorb most of the manager's discount. Those flexible-wage outcomes require a floor of at most 0.41, which this economy does not have; with the floor of 0.60, patient firms run the pipeline down to two seniors, and the response to AI lowers welfare by 7.2, more than the entire first-best gain. Myopia then compounds: at $\beta^M = 0.90$ the stock falls to 1.34, a quarter of the pre-AI level and just over a third of the first best, the senior wage rises to nearly three times its pre-AI level, and the welfare loss is 22.5, four times the first-best gain. A decision-maker with $\beta^M = 0.80$ rather than 0.90 runs the pipeline down to 0.68 and the loss to 44.7.

Credit reform is the most effective remedy in this calibration. Raising conventional credit from 0.10 to 0.29 makes the floor slack at the flexible myopic stock and moves the economy from the −22.45 row to the +3.87 row, with no residency and no resource cost. The residency rows show what the intermediary can do where that reform is unavailable. The unsubsidized residency collects a small net tuition of 0.05, maintains a stock of 2.08, and recovers about forty percent of the myopic loss, stopping well short of the flexible-wage market because it collects none of the firms' rent. A subsidy equal to the supervision cost raises the stock

to 2.56. The subsidy that implements the first-best stock, 0.61, achieves it at a welfare of -11.08, below the residency-efficient outcome of -8.69 at the smaller subsidy 0.43, because supervision is a real cost. Even that outcome is below the pre-AI economy: in this calibration supervision costs exceed the price of the machine, so the residency limits the loss from AI rather than restoring the old ladder. The residency rows do not depend on the manager's horizon, since the intermediary's financing does not, but they depend on graduates' wealth and credit, as Table 2 shows.

The welfare cutoff is $\bar{\kappa} = 0.4675$ and the activity cutoff is $\kappa^{\text{act}} = 0.9159$. At $\kappa = 0.57$, for example, the program is privately feasible and active, with a stock of 1.65, but lowers welfare by 3.44 relative to the myopic market: private financing and social desirability are different conditions. Capacity matters as well. With the baseline parameters the market's replacement flow is $\delta S^M/m = 1.34$ per firm. A program with capacity $\bar{J}^R = 1$ leaves the stock at 1.34 and wastes 0.25 of supervision per firm per period; capacity 2 raises it to 2.00; and only capacity of 2.08 or more delivers the full $S^R = 2.08$. Scale-dependent supervision lowers the stock as well: at $\kappa_1 = 0.05$, 0.10, and 0.20 the steady-state residency stock is 1.93, 1.81, and 1.64, against 2.08 at $\kappa_1 = 0$ and a market stock of 1.34.

Table 2: Graduates' wealth, the floor, and the residency.

Assets a	Floor $\underline{w}$	Market stock S^M	Market welfare $W^M - W^{\text{pre}}$	Residency active?	Residency welfare $W^R - W^{\text{pre}}$	Credit needed $b_U \geq$
0.0	0.70	0.99	-32.81	yes	-13.76	0.85
0.1	0.60	1.34	-22.45	yes	-13.76	0.75
0.2	0.50	1.96	-8.88	yes‡	-13.76	0.65
0.3	0.40	3.16	$+3.87$	no	—	0.55

Note: $\ell = 0.8$, $b_F = 0.1$, $b_U = 1.5$; other parameters as in Table 1; $g = 0$. With the floor at 0.4 the market clears at a wage of 0.41 and the floor is slack. ‡Active but welfare-reducing: $W^R < W^M$. The last column is the credit needed for condition (20).

Wealth and credit are the comparative statics of the liquidity interpretation. As graduates' assets rise, the floor falls, the market pipeline recovers, and the residency's advantage shrinks: it stops being welfare-improving at a floor of about 0.53 and stops being active at about 0.49 (assets of about 0.17 and 0.21), and where the floor is slack the market outperforms it. Conversely the residency requires student credit that covers living costs and net tuition, at least 0.75 for the baseline, and is infeasible for less credit.

6. Implications for University Education

The model's prescription for universities is more specific than the injunction to teach judgment. Four implications follow from the structure of Proposition 10.

The first concerns design. The intermediary's advantage rests on credible training, Assumption 4(b). In practice this means programs in which producing the deliverable requires the exercise of judgment under supervision, with process records that make the training verifiable: embedded consulting projects under faculty direction, clinical placements, and structured engagements in which student teams take a sponsoring firm's live problem from an ambiguous brief to a recommendation. The condition is not that the program be experiential but that the experience be the one AI removed. A task can teach a human even if a machine could produce its final output equally well, which is the learning-by-doing mechanism; the concern is a program that submits an AI-produced deliverable while bypassing the trainee's developmental work, and a firm sponsoring such a program is buying routine work and not producing seniors. Certification must distinguish those processes, at a cost, and university ownership alone cannot. Medicine solved this problem long ago—structured supervision, progressive responsibility, and real patients as the training substrate—and the case method, clinical legal education, and consulting practica are existing models in business and law. Generative AI itself can provide part of the substrate: simulated negotiations with a machine counterpart, deal processes in which students must find the assumptions in a machine-built model that require human judgment, litigation exercises in which students review machine-generated document sets. Simulation lowers the deliverable's value and the fidelity of the training, which the model registers as a higher net cost and a lower conversion probability, but it also lowers the cost of repetition on which the Dreyfus and Dreyfus (1986) account of expertise rests.

The second concerns scale. Proposition 10(iv) says that a program smaller than the market's replacement flow displaces firm training one for one, and Proposition 11 says that supervision costs rise with volume, so a program that moves the profession's stock must be sized to the pipeline and priced at its marginal cost. Existing consulting practica are small relative to the entry-level cohorts that a management consulting firm once absorbed, and the model implies that their aggregate effect on the pipeline is correspondingly small. A program is viable where the tuition graduates would pay at the market steady state covers its net cost, the condition of Proposition 10(ii): where the floor is high relative to the machine and seniors are scarce, and not where graduates are wealthy or credit is ample.

The third concerns finance. The binding constraint in the model is liquidity, and the

university helps because it can obtain student credit for the graduate. Extending credit to conventional apprenticeships at accredited employers would relax the same constraint directly and without a supervision cost, and where it is feasible it is the first reform to consider; the intermediary matters where enforceability limits conventional credit but not educational credit, or where public funds are costly. The model's prediction that junior hiring in AI-exposed occupations falls most where entry-level workers have the least wealth and credit can be tested directly.

The fourth concerns public funding. The intermediary cannot collect the rent that trained seniors generate for whichever firm employs them, so implementing the efficient stock requires a subsidy, but the subsidy that does so is $g^{*,R} = \beta mv(S^{*,R})$ and not $\kappa+\beta mv(S^*)$: supervision is a real cost that a transfer does not remove, and it lowers the efficient stock. This is the economic logic of public support for graduate medical education, and the model extends it to the knowledge-work professions in which AI has removed the market's own financing of apprenticeship. When public funds are costly, an unsubsidized program that raises the stock with no outlay has a distinct advantage over verified direct subsidies (Proposition 12). The alternative instruments of Section 4 do not substitute for the subsidy at the efficient scale: a deferred rebate must be scaled to the manager's horizon, and a hiring subsidy is exposed to diversion. The model supplies the calibration target: $g^{*,R}$ can be estimated from the firm's share of senior rents and the maturation probability.

7. Conclusion

Generative AI makes routine cognitive work cheap. The paper's argument is that the consequences for the professions run through a price rather than a capability: the machine does not change how judgment is formed, but it changes what a firm must pay to have a junior do the work through which judgment was formed. Before AI, the junior was essential and the training was free. After AI, the junior costs a premium over the machine, and the premium must be financed by someone.

Who finances it determines the outcome. Workers would, and where compensation is flexible they do: the senior premium rises as the pipeline thins, graduates accept lower entry wages to buy the training, and neither poaching nor managerial myopia does much harm. A liquidity constraint—living costs net of graduates' wealth and credit, together with legal limits on sub-minimum and unpaid work—blocks this channel, and in the model the channel is exactly the one AI pushes against, because the wage graduates would accept falls when seniors become scarce. Firms then finance training they cannot sell, and a decision-maker

with a short horizon finances little. The senior stock settles below its efficient level and the senior wage premium above it; the pipeline does not vanish, but it can shrink enough that the market's response to AI is worse than never adopting it.

The remedies are ranked by what they require. Extending credit to conventional apprenticeships restores flexible compensation at no resource cost, and is the first reform to consider. Verified employer subsidies can implement the efficient stock, but they require verification of developmental content and public funds, and deferred payments must scale with the inverse of the manager's horizon. A training intermediary that can charge tuition financed by student credit, and can credibly certify what it teaches, avoids both the need for conventional credit and the need for public funds, not because it is patient but because it can be paid by the party who is. It is a second-best institution: its supervision costs are real, so even the subsidy that implements the original efficient stock does not restore first-best welfare; it must be large enough to exceed the market's replacement flow to raise the stock at all; it can be active and yet reduce welfare; it depends on graduates' credit and on a supply of supervised projects; and it is dominated by a market in which compensation is flexible. The broken ladder is not a temporary disruption but a financing problem created by a price change, and the task the paper sets for higher education is to build the institutions that can finance and credibly deliver apprenticeship where the market no longer does.

References

Acemoglu, D. (2024). The simple macroeconomics of AI. NBER Working Paper No. 32487.

Acemoglu, D. and D. H. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4:1043–1171.

Acemoglu, D., D. H. Autor, and S. Johnson (2026). Building pro-worker artificial intelligence. NBER Working Paper No. 34854.

Acemoglu, D., D. Kong, and A. Ozdaglar (2026). AI, human cognition and knowledge collapse. NBER Working Paper No. 34910.

Acemoglu, D. and J. S. Pischke (1998). Why do firms train? Theory and evidence. *Quarterly Journal of Economics*, 113(1):79–119.

Acemoglu, D. and J. S. Pischke (1999). Beyond Becker: Training in imperfect labour markets. *Economic Journal*, 109(453):112–142.

Acemoglu, D. and P. Restrepo (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Acemoglu, D. and P. Restrepo (2022). Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, 90(5):1973–2016.

Aghion, P., B. F. Jones, and C. I. Jones (2018). Artificial intelligence and economic growth. In A. K. Agrawal, J. Gans, and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, pp. 237–282. University of Chicago Press.

Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29(3):155–173.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024). New frontiers: The origins and content of new work, 1940–2018. *Quarterly Journal of Economics*, 139(3):1399–1465.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2026). What makes new work different from more work? *NBER Working Paper*, 34986.

Barron, J. M., M. C. Berger, and D. A. Black (1999). Do workers pay for on-the-job training? *Journal of Human Resources*, 34(2):235–252.

Becker, G. S. (1964). *Human Capital*. University of Chicago Press.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365.

Biasi, B., and S. Ma (2022). The education-innovation gap. NBER Working Paper No. 29853.

Biasi, B., and S. Ma (2026). Frontier knowledge in college and student success. Working Paper.

Brynjolfsson, E., B. Chandar, and R. Chen (2025). Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence. Stanford Digital Economy Lab Working Paper.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at work. *Quarterly Journal of Economics*, 140(2):889–942.

Cunha, F. and J. J. Heckman (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Dell'Acqua, F., E. McFowland III, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Working Paper No. 24-013.

Dreyfus, H. L. and S. E. Dreyfus (1986). *Mind Over Machine*. Free Press.

Edmondson, A. C. and T. Chamorro-Premuzic (2025). The perils of using AI to replace entry-level jobs. *Harvard Business Review*, September 2025.

Eisfeldt, A. L. and D. Papanikolaou (2013). Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4):1365–1406.

Eisfeldt, A. L., G. Schubert, and M. B. Zhang (2023). Generative AI and firm values. NBER Working Paper No. 31222.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, 384(6702):1306–1308.

Federal Student Aid, U.S. Department of Education (n.d.). Eligibility for federal student aid. `https://studentaid.gov/understand-aid/eligibility/requirements`. Accessed September 28, 2026.

Fuller, J., M. Sigelman, and M. Fenlon (2025). How gen AI could change the value of expertise. *Harvard Business Review*, March 2025.

Gans, J. S. (2025). *The Microeconomics of Artificial Intelligence*. MIT Press.

Gans, J. S. (2025). Growth in AI knowledge. NBER Working Paper No. 33907.

Gao, J., W. Wang, and Y. Wu (2024). Human capital portability and careers in finance. *Review of Financial Studies*, 37(9):2732–2778.

Gathmann, C. and U. Schönberg (2010). How general is human capital? A task-based approach. *Journal of Labor Economics*, 28(1):1–49.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40(1–3):3–73.

Heckman, J. J. (1998). What should be our human capital investment policy? *Fiscal Studies*, 19(2):103–119.

Hosseini, S. M. and G. Lichtinger (2025). Generative AI as seniority-biased technological change. SSRN Working Paper No. 5425555.

Jackson, M. O. and T. R. Palfrey (2001). Voluntary implementation. *Journal of Economic Theory*, 98(1):1–25.

Jones, C. I. (2026). A.I. and our economic future. NBER Working Paper No. 34779.

Loewenstein, M. A. and J. R. Spletzer (1998). Dividing the costs and returns to general training. *Journal of Labor Economics*, 16(1):142–171.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press.

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics*, 13(4):653–677.

Noy, S. and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint* arXiv:2302.06590.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Spence, M. and D. Starrett (1975). Most rapid approach paths in accumulation problems. *International Economic Review*, 16(2):388–403.

Stein, J. C. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *Quarterly Journal of Economics*, 104(4):655–669.

U.S. Department of Labor, Wage and Hour Division (2018). Fact Sheet #71: Internship programs under the Fair Labor Standards Act (updated January 2018). `https://www.dol.gov/agencies/whd/fact-sheets/71-flsa-internships`. Accessed September 28, 2026.

Walther, C. C. (2025). Is AI pushing us to break the talent pipeline? Knowledge@Wharton, August 2025.

Zingales, L. (2000). In search of new foundations. *Journal of Finance*, 55(4):1623–1653.

Internet Appendix for

The Broken Job Ladder: AI, Vanishing Apprenticeships, and the Role of Universities

Miao Ben Zhang

A. Proofs

Throughout, $r \equiv 1-\beta(1-\delta)$, $r^M \equiv 1-\beta^M(1-\delta)$, and $\bar{a} \equiv 1-\delta$ is the retention factor. Then $\Delta(S) = [\theta G'(S) - \omega]/r$, $v(S) = (1-\theta)G'(S)/r$, and $v^M(S) = (1-\theta)G'(S)/r^M$. Since $\beta^M \leq \beta$, $r^M \geq r$ and $v^M(S) \leq v(S)$. The premium function $p(S) = \max\{\underline{w}, \omega - \beta m \Delta(S)\} - c$ is weakly increasing in S.

A.1 Worker Financing and Participation

A loan d has entry-date repayment value $\sum_{u=1}^{\infty} \beta^u \frac{1-\beta}{\beta} d = d$, so it contributes zero to net present value while permitting current consumption. The consumption inequality $w^J + a + d \geq \ell$ with $d \leq b_F$, together with the legal wage rule, gives $\underline{w}$. The participation comparison between a one-period apprenticeship and permanent outside work is $w^J - \omega + \beta m D_{t+1} \geq 0$: initial assets cancel and repayments exactly offset borrowing. Excess supply of eligible graduates then gives (6). A resident instead has first-period earnings $-\tau$, so participation is $-\tau - \omega + \beta m D_{t+1} \geq 0$, and current financing requires $\tau + \ell \leq a + b_U$. Condition (5) ensures that the annuity $(1-\beta)d/\beta$ with $d \leq b_U$ is at most $\omega - \ell$, the income net of subsistence that every worker can earn in later periods, so repayment respects subsistence even after failure. These arguments also show why a large lifetime return alone does not relax a current borrowing limit.

A.2 Proof of Lemma 1

For given wage and matching sequences the period payoff is concave in the state and affine in hiring, the law of motion is affine, and the choice interval is convex. The first-order hiring condition is $-p_t + \beta^M(1-\lambda)mV'_{t+1}(S_{i,t+1})$, and the envelope equation is $V'_t(S_{i,t}) = G'(S_{i,t}) - w^S_t + \beta^M \bar{a} V'_{t+1}(S_{i,t+1})$, which at symmetry yields the recursion for Q_t in (3). For a direct verification of sufficiency, consider any feasible deviation with state difference $\Delta S_{i,t}$

and hiring difference $\Delta J_{i,t}$ from a proposed symmetric path. Concavity gives

$$\Delta\pi_{i,t} \;\le\; \sigma(S_t)\,\Delta S_{i,t} - p_t\,\Delta J_{i,t}, \qquad \Delta S_{i,t+1} = \bar{a}\Delta S_{i,t} + (1-\lambda)m\Delta J_{i,t}.$$

Using $Q_t - \beta^M \bar{a} Q_{t+1} = \sigma(S_t)$ and telescoping, the discounted payoff difference is at most

$$\sum_{t\ge 0} (\beta^M)^t \big[\beta^M (1-\lambda) m Q_{t+1} - p_t\big]\Delta J_{i,t}.$$

The initial state difference is zero and the terminal term vanishes under bounded values and states. Each summand is nonpositive under the stated conditions: if $J_{i,t} = 0$ then $\Delta J_{i,t} \ge 0$ and the bracket is nonpositive; if $J_{i,t} = \bar{J}$ then $\Delta J_{i,t} \le 0$ and the bracket is nonnegative; and the bracket is zero at an interior choice. This verifies global optimality rather than treating a steady-state value as a transition value. The same argument applies with an exogenous residency allocation and the corresponding residual hiring capacity. □

A.3 Proof of Proposition 1

With $c = \infty$ the routine requirement can be met only by juniors, so $J = \bar{J}$ regardless of the value of maturation, and the steady state of $S_{t+1} = (1-\delta)S_t + m\bar{J}$ is $S^{\text{pre}} = m\bar{J}/\delta$. Neither λ nor β^M enters. The junior wage follows from (6) at $D_{t+1} = \Delta(S^{\text{pre}})$. □

A.4 Proof of Proposition 2

Define $\Lambda(S) \equiv \beta^M(1-\lambda)m v^M(S) - p(S)$. The first term is strictly decreasing in S with $\lim_{S\to 0} v^M(S) = \infty$ and $\lim_{S\to\infty} v^M(S) = 0$; the second is weakly increasing, with $\lim_{S\to\infty} p(S) = \omega(1+\beta m/r) - c > 0$ since $\Delta \to -\omega/r$. Hence Λ is continuous and strictly decreasing, positive near zero and negative for large S, with a unique root $S^e > 0$. This proves the preamble and part (iv). A steady state has $J = \delta S/m$, feasible if and only if $S \le S^{\text{pre}}$.

Part (i). $\Lambda(S^{\text{pre}}) \ge 0$ if and only if $S^e \ge S^{\text{pre}}$, and $\Lambda(S^{\text{pre}}) \ge 0$ rearranges to $c \ge \hat{c}$. On the constant path $D_t = \Delta(S^{\text{pre}})$ and $Q_t = v^M(S^{\text{pre}})$, so $p_t - \beta^M(1-\lambda)mQ_{t+1} = -\Lambda(S^{\text{pre}}) \le 0$, and $J_t = \bar{J}$ satisfies Lemma 1.

Part (ii). Let $c < \hat{c}$, so $S^e < S^{\text{pre}}$. Write the path as $S_t = \max\{\bar{a}^t S^{\text{pre}}, S^e\}$ and $J_t = (S_{t+1} - \bar{a}S_t)/m$, and let T be the first date with $S_T = S^e$. The path is feasible: $J_t = 0$ before the crossing date, $J_t = \delta S^e/m \le \bar{J}$ afterward, and in the crossing period $J_t = (S^e - \bar{a}S_t)/m < \delta S^e/m$ because $S_t > S^e$. Along the path, $S_{t+s} \ge S^e$ for all s, and G' is

decreasing, so

$$D_t = \sum_{s\geq 0}(\beta\bar{a})^s[\theta G'(S_{t+s}) - \omega] \leq \Delta(S^e), \qquad Q_t = \sum_{s\geq 0}(\beta^M\bar{a})^s\sigma(S_{t+s}) \leq v^M(S^e),$$

with equality if and only if $t \geq T$. Assumption 1(c) ensures $\theta G'(S_{t+s}) \geq \omega$ on the path. *Dates* $t \geq T-1$ (the crossing period and after): $t+1 \geq T$, so $D_{t+1} = \Delta(S^e)$ and $Q_{t+1} = v^M(S^e)$, and $p_t = p(S^e) = \beta^M(1-\lambda)mv^M(S^e)$: the interior condition holds with equality. *Dates* $t < T-1$*:* $J_t = 0$ and $D_{t+1} \leq \Delta(S^e)$, hence $\omega - \beta m D_{t+1} \geq \omega - \beta m\Delta(S^e)$ and $p_t \geq p(S^e) = \beta^M(1-\lambda)mv^M(S^e) \geq \beta^M(1-\lambda)mQ_{t+1}$, so the condition for $J_t = 0$ holds.

Part (iii). Λ is decreasing in S and increasing in c, so S^e is increasing in c. Since $S^e < S^{\text{pre}}$ and G' is decreasing, $w^S = \theta G'(S^e)$ exceeds its pre-AI level and $\Delta(S^e) > \Delta(S^{\text{pre}})$, so $\omega - \beta m\Delta(S^e) < \omega - \beta m\Delta(S^{\text{pre}})$. □

A.5 Proof of Proposition 3

Write the period return as $F(S_t) - (\omega - c)J_t$ with $F(S) \equiv G(S) - \omega S - c\bar{J}$, and substitute $J_t = [S_{t+1} - (1-\delta)S_t]/m$. Using $\sum_{t\geq 0}\beta^t S_{t+1} = \beta^{-1}[\sum_{t\geq 0}\beta^t S_t - S_0]$,

$$\begin{aligned}\sum_{t\geq 0}\beta^t\big[F(S_t) - (\omega-c)J_t\big] &= \sum_{t\geq 0}\beta^t\Big[F(S_t) - (\omega-c)\frac{S_t}{\beta m} + (\omega-c)\frac{(1-\delta)S_t}{m}\Big] + \frac{(\omega-c)S_0}{\beta m} \\ &= \sum_{t\geq 0}\beta^t\Phi(S_t) + \frac{(\omega-c)S_0}{\beta m},\end{aligned}$$

with Φ as in (11). Φ is strictly concave, with $\Phi'(S) = G'(S) - \omega - (\omega-c)r/(\beta m)$, which is zero at the unique S^* satisfying $G'(S^*) - \omega = (\omega-c)r/(\beta m)$; using $\Delta + v = (G'-\omega)/r$ this is (12). Feasible paths satisfy $(1-\delta)S_t \leq S_{t+1} \leq (1-\delta)S_t + m\bar{J}$, so $S_t \geq (1-\delta)^t S_0$ and $S_t \leq S^{\text{pre}}$ when $S_0 = S^{\text{pre}}$. If $S^* < S^{\text{pre}}$, let $S_t^* = \max\{(1-\delta)^t S_0, S^*\}$. If $S_t \geq S^*$ then $S_t \geq S_t^* \geq S^*$ and Φ is decreasing above S^*, so $\Phi(S_t) \leq \Phi(S_t^*)$; if $S_t < S^*$ then $(1-\delta)^t S_0 \leq S_t < S^*$ so $S_t^* = S^*$ and $\Phi(S_t) \leq \Phi(S^*)$. Hence the most rapid approach maximizes every term. If $S^* \geq S^{\text{pre}}$, Φ is increasing on $[0, S^{\text{pre}}]$ and the stock is maximized by $J_t = \bar{J}$. Finally, $S^* < S^{\text{pre}}$ if and only if $\Phi'(S^{\text{pre}}) < 0$, which rearranges to $c < c^*$; and

$$\begin{aligned}\hat{c} - c^* &= \big[\max\{\underline{w}, \omega - \beta m\Delta(S^{\text{pre}})\} - (\omega - \beta m\Delta(S^{\text{pre}}))\big] \\ &\qquad + \big[\beta m v(S^{\text{pre}}) - \beta^M(1-\lambda)mv^M(S^{\text{pre}})\big] \;\geq\; 0,\end{aligned}$$

because the first bracket is nonnegative by definition of the maximum and the second is nonnegative since $\beta \geq \beta^M$, $v \geq v^M$, and $\lambda \geq 0$, and it is strictly positive if $\lambda > 0$ or $\beta^M < \beta$. □

A.6 Proof of Proposition 4

With the floor slack and $\beta^M = \beta$, (9) is $\omega - c = H^B(S) \equiv \beta m[\Delta(S) + (1-\lambda)v(S)]$, while (12) is $\omega - c = H^*(S) \equiv \beta m[\Delta(S) + v(S)]$. Both are strictly decreasing with the Inada limits, and $H^B \leq H^*$ pointwise with equality if and only if $\lambda = 0$. Hence $S^B \leq S^*$ with equality if and only if $\lambda = 0$. Writing $H^B(S) = \beta m\{[1 - \lambda(1-\theta)]G'(S) - \omega\}/r$ shows that H^B is decreasing in λ and increasing in θ pointwise and converges to H^* as $\theta \to 1$, so S^B falls with λ, rises with θ, and converges to S^*. With a myopic decision-maker and flexible wages, (9) is $\omega - c = \beta m \Delta(S) + \beta^M(1-\lambda)mv^M(S)$; the right side weakly exceeds $\beta m\Delta(S)$ for every S and every $\beta^M \geq 0$, so its root satisfies $\beta m\Delta(S) \leq \omega - c$, that is $S \geq \tilde{S}$, where $\tilde{S} > 0$ by the Inada condition. The right side is strictly increasing in β^M for each S (both β^M and $1/r^M$ rise), so the root $S^{B,M}$ is increasing in β^M; removing the firm's term as $\beta^M \downarrow 0$ gives $S^{B,M} \to \tilde{S}$ by continuity and strict monotonicity. With $G'(S) = AS^{-\eta}$, $\beta m\Delta(\tilde{S}) = \omega - c$ and $H^*(S^*) = \omega - c$ give $\theta A\tilde{S}^{-\eta} = A\,S^{*-\eta}$, so $\tilde{S}/S^* = \theta^{1/\eta}$. □

A.7 Proof of Proposition 5

Part (i): at $\hat{S}$ the two branches of (6) coincide, so $p(\hat{S}) = \underline{w} - c$. The steady state S^e lies below $\hat{S}$ if and only if $\Lambda(\hat{S}) < 0$, that is $\beta^M(1-\lambda)mv^M(\hat{S}) < \underline{w} - c$; since the left side is positive, this requires $\underline{w} > c$. Before AI the stock is S^{pre}, and the floor is slack at S^{pre} if and only if $S^{\text{pre}} > \hat{S}$ (or equal). Part (ii): with the floor binding, (9) at $\beta^M = \beta$ is (15), whose right side is strictly decreasing in S and whose left side is a constant increasing in $\underline{w}$; hence S^F is unique and decreasing in $\underline{w}$, and Δ does not appear. If the floor binds at S^B, then $\omega - \beta m\Delta(S^B) < \underline{w}$, so at S^B the left side of (15) exceeds $\omega - c - \beta m\Delta(S^B) = \beta(1-\lambda)mv(S^B)$, the right side; since the right side is decreasing, $S^F < S^B$. Part (iii) follows from (ii) and $\underline{w} = \ell - a - b_F$. □

A.8 Proof of Proposition 6

The right side of (16), $\beta^M(1-\lambda)m(1-\theta)G'(S)/r^M$, is strictly decreasing in S, strictly increasing in β^M for each S (both β^M and $1/r^M$ rise), and lies strictly below $\beta(1-\lambda)mv(S)$ when $\beta^M < \beta$; hence S^M is unique, increasing in β^M, and below S^F. If $\underline{w} > c$, then $\underline{w} - c > \beta^M(1-\lambda)mv^M(\hat{S})$ for sufficiently small β^M, so the floor binds by Proposition 5(i).

As $\beta^M \downarrow 0$ the right side of (16) converges to zero for every $S > 0$, so the root converges to zero; for every $\beta^M > 0$ the root is positive because $v^M \to \infty$ as $S \to 0$. With $G'(S) = AS^{-\eta}$, solving (16) and the defining equation of S_0^F for $S^{-\eta}$ and taking the ratio gives $(S^M/S_0^F)^{-\eta} = \beta r^M/[\beta^M(1-\lambda)r]$, which is the stated expression. □

A.9 Proof of Proposition 7

By Proposition 3, $W(s) = N\Psi(s) + \text{const}$ with $\Psi(s) \equiv \sum_t \beta^t \Phi\big(\max\{(1-\delta)^t S^{\text{pre}}, s\}\big)$ and a constant independent of the regime. Part (i): for $s_a < s_b \le S^*$, at each t either $\max\{(1-\delta)^t S^{\text{pre}}, s_a\} = \max\{(1-\delta)^t S^{\text{pre}}, s_b\}$, or $s_a \le \max\{(1-\delta)^t S^{\text{pre}}, s_a\} < \max\{(1-\delta)^t S^{\text{pre}}, s_b\} \le S^*$, where Φ is strictly increasing; the second case obtains for all large t, so $\Psi(s_a) < \Psi(s_b)$. The stock comparisons $S^M \le S^F \le S^B \le S^*$ from Propositions 4–6 then give the chain, with the stated strictness conditions. Part (ii): the pre-AI path is the constant path at S^{pre}, which is feasible after AI, so $W^* \ge W^{\text{pre}}$; the inequality is strict if and only if $S^* < S^{\text{pre}}$, that is $c < c^*$ (Proposition 3). Part (iii): all stocks lie in $[0, S^{\text{pre}}]$ and G is continuous at zero, so discounted flow returns are bounded by a summable sequence and dominated convergence makes $W(s)$ continuous on $[0, S^*]$. Under Assumption 3, $W(S^*) > W^{\text{pre}}$ by part (ii). If $W(0) < W^{\text{pre}}$, the intermediate value theorem and strict monotonicity give a unique crossing $\underline{S} \in (0, S^*)$. □

A.10 Proof of Proposition 8

If $w_{\min} \le w^{B,M}$ and $b_F \ge \ell - a - w^{B,M}$ then $\underline{w} = \max\{w_{\min}, \ell - a - b_F\} \le w^{B,M}$. At $S^{B,M}$ the market-clearing wage $\omega - \beta m \Delta(S^{B,M}) = w^{B,M}$ is then at least the floor, so $p(S^{B,M})$ equals its flexible-compensation value and $S^{B,M}$ satisfies (9); by uniqueness of the root (Proposition 2) it is the steady state, and the transition follows from the same proposition. Credit is fairly priced and fully recourse under Assumption 2, so the reform has no resource or fiscal cost. The patient case replaces β^M by β. □

A.11 Proof of Proposition 9

Part (i): with a payment ρ per separated matured junior, the decision-maker's return on a junior is $\beta^M m[(1-\lambda)v^M(S) + \lambda\rho]$, which at $\rho = v^M(S)$ equals $\beta^M m v^M(S)$; the steady-state condition becomes $p(S) = \beta^M m v^M(S)$. With flexible wages and $\beta^M = \beta$ this is $\omega - c - \beta m \Delta(S) = \beta m v(S)$, which is (12). When the floor binds at S^*, $p(S^*) = \underline{w} - c > \omega - c - \beta m \Delta(S^*) = \beta m v(S^*)$, so the root lies below S^*; when $\beta^M < \beta$ the factor $\beta^M v^M$ is untouched. Part (ii): a verified current subsidy replaces the premium by $p(S) - s$. At

S^*, $p(S^*) \geq \omega - \beta m \Delta(S^*) - c = \beta m v(S^*) \geq \beta^M (1-\lambda) m v^M(S^*)$, so $s^* \geq 0$, and choosing (17) makes the interior stationary condition hold at S^*. The modified stationary equation is strictly increasing in S, so the target is unique, and the coast-and-maintain path is verified as in Proposition 2 with the premium reduced by the constant s^*. Part (iii): a next-period payment r per qualification adds $\beta^M m r$ to the hiring return; equating it to the required current subsidy gives $\beta^M m r^* = s^*$, which is (18). As $\beta^M \downarrow 0$, $p(S^*) > 0$ remains fixed and $v^M(S^*)$ remains finite, so r^* diverges; it is decreasing in β^M because $p(S^*)/(\beta^M m)$ falls and $(1-\lambda)v^M(S^*)$ rises. This proves the incentive results without any claim about voluntary participation. □

A.12 Proof of Proposition 10

Part (i). The firm's constraint is $f_t \leq c$ and the resident's are $\tau_t \leq \beta m D_{t+1} - \omega$ and $\tau_t \leq a + b_U - \ell$; none contains β^M, and the actions of other firms enter only through the stock.

Part (ii). Eliminating the fee and tuition, a slot is feasible exactly when (20) and (21) hold: necessity follows from $f_t \leq c$, break-even, and the two resident constraints, and sufficiency follows by taking $f_t = c$ and $\tau_t = \kappa - c - g$. At a constant stock (21) is $\beta m \Delta(S) \geq \omega - c + \kappa - g$, and since Δ is decreasing this holds if and only if $S \leq S^R$. Hence $S^R > S^M$ if and only if $\beta m \Delta(S^M) > \omega - c + \kappa - g$, which rearranges to (25). For the explicit form in the text, (16) gives $G'(S^M) = (\underline{w} - c) r^M / [\beta^M (1-\lambda) m (1-\theta)]$, so $\beta m \Delta(S^M) = (\underline{w} - c)\, \theta \beta r^M / [(1-\theta)\beta^M (1-\lambda) r] - \beta m \omega / r$, and the inequality follows.

Part (iii). From S^{pre} the residency is inactive while $S_{t+1} > S^R$: then $D_{t+1} < \Delta(S^R)$ makes (21) fail (as in the proof of Proposition 2). In the crossing period $S_{t+1} = S^R$ and $D_{t+1} = \Delta(S^R)$, so (21) holds with equality and the residency admits $(S^R - (1-\delta)S_t)/m < \delta S^R/m \leq \bar{J}$ residents, which capacity permits. Firms never hire: along the path $Q_{t+1} \leq v^M(S^R) < v^M(S^M)$ since $S^R > S^M$, and $p_t \geq p(S^R) \geq p(S^M) = \beta^M (1-\lambda) m v^M(S^M) > \beta^M (1-\lambda) m Q_{t+1}$, using that p is weakly increasing, so Lemma 1 gives $J_t = 0$. The residency then supplies the whole flow $\delta S^R/m$ at net tuition τ^0.

Part (iv). Let $S^R > S^M$ and (20) hold. (a) If $\bar{J}^R \leq \delta S^M/m$, students' surplus at S^M is nonnegative, so the provider fills capacity, and firms hire the balance $\delta S^M/m - \bar{J}^R \geq 0$, at which the firm's interior condition holds with equality; the stock is S^M. (b) If $\delta S^M/m \leq \bar{J}^R \leq \delta S^R/m$, the residency at capacity sustains $S = m\bar{J}^R/\delta \in [S^M, S^R]$, at which $\Lambda(S) \leq 0$ so firms hire nothing, and (21) holds since $S \leq S^R$. (c) If $\bar{J}^R \geq \delta S^R/m$, part (iii) applies. These cases give (26)–(27); at the boundaries the formulas agree, and $J + J^R = \delta S/m \leq \bar{J}$ throughout.

Part (v). Since $\Delta+v$ is decreasing, $S^{*,R} < S^*$ because the left side of (23) exceeds that of (12). $S^R < S^*$ if and only if $\beta m\Delta(S^*) < \omega-c+\kappa-g$; by (12), $\beta m\Delta(S^*) = \omega-c-\beta mv(S^*)$, so the condition is $g < \kappa+\beta mv(S^*)$. By (23), $\beta m\Delta(S^{*,R}) = \omega-c+\kappa-\beta mv(S^{*,R})$, so $S^R < S^{*,R}$ if and only if $g < \beta mv(S^{*,R}) = g^{*,R}$. By (13), $S^R < S^B$ if and only if $g < \kappa+\beta(1-\lambda)mv(S^B)$.

Part (vi). On the same stock path, welfare in the residency regime is the planner's welfare less $N\kappa L(s)$, since supervision resources are consumed and transfers cancel; this is strictly below W^* at $s = S^*$ if $\kappa > 0$. The planner constrained to the residency technology solves (10) with $\omega - c + \kappa$ in place of $\omega - c$, which by the argument of Proposition 3 has target $S^{*,R}$; the subsidy $g^{*,R}$ implements it by part (v) and (22), provided (20) and capacity do not bind. When direct training remains available, the residency is active only if $S^{*,R} > S^M$ by part (ii), and otherwise firms would train at S^M.

Part (vii). Let Φ_κ be Φ with $\omega-c+\kappa$ in the linear term and $\Psi_\kappa(s) \equiv W(s) - N\kappa L(s) = N\sum_t \beta^t \Phi_\kappa(S_t(s))+\text{const}_\kappa$, so that $W^R(\kappa) = \Psi_\kappa(S^R(\kappa))$. By the argument of Proposition 7(i) applied to Φ_κ, Ψ_κ is strictly increasing on $[0, S^{*,R}(\kappa)]$, and $S^R(\kappa) < S^{*,R}(\kappa)$ when $g = 0$. For $\kappa_1 < \kappa_2$ in the active range, $S^R(\kappa_2) < S^R(\kappa_1)$ since Δ is decreasing, and $\Psi_{\kappa_2}(s) = \Psi_{\kappa_1}(s) - (\kappa_2 - \kappa_1)NL(s) < \Psi_{\kappa_1}(s)$ because $L(s) > 0$ for $s < S^{\text{pre}}$. Therefore $W^R(\kappa_2) < \Psi_{\kappa_1}(S^R(\kappa_2)) < \Psi_{\kappa_1}(S^R(\kappa_1)) = W^R(\kappa_1)$, and continuity follows from continuity of S^R and of the discounted sums. The condition $\kappa^{\text{act}} > 0$ is equivalent to $S^R(0) > S^M$; with $S^R(0) < S^*$, Proposition 7(i) gives $W^R(0) = W(S^R(0)) > W(S^M) = W^M$. As $\kappa \uparrow \kappa^{\text{act}}$, $S^R(\kappa) \downarrow S^M$ and $W^R(\kappa) \to W(S^M) - N\kappa^{\text{act}} L(S^M) < W^M$. A strictly decreasing continuous function that crosses W^M does so once, giving $\bar{\kappa} \in (0, \kappa^{\text{act}})$. Inactivation changes the training technology, so continuity across it is neither required nor used. □

A.13 Proof of Proposition 11

At a steady state, $J^R = \delta S/m$ per firm, so the break-even condition of an active residency is $\beta m\Delta(S) = \omega - c + \kappa_0 + \kappa_1\delta S/m - g$. The left side is strictly decreasing in S and the right side strictly increasing, so there is a unique root, and implicit differentiation shows that it is decreasing in κ_1 and κ_0. The residency is active if and only if the root exceeds S^M (by the argument of Proposition 10(ii), because at $S \le S^M$ firms hire) and (20) holds with κ replaced by $\kappa(\delta S^R/m)$. □

A.14 Proof of Proposition 12

For $S > S^M$, $\Lambda(S) < 0$ because Λ is strictly decreasing with $\Lambda(S^M) = 0$, that is $p(S)-\beta^M(1-\lambda)mv^M(S) > 0$. A verified direct subsidy must cover the verification cost κ_F that enters the

firm's premium and close this gap, so $s(S) = \kappa_F + p(S) - \beta^M(1-\lambda)mv^M(S) > \kappa_F \geq 0$ per trainee, financed at cost $1+\phi$; the flow of trainees at stock S is $\delta S/m$ per firm. For the residency, by Proposition 10(iii)–(iv) with $g=0$ the stock $S \in (S^M, S^R]$ is sustained by capacity $\bar{J}^R = \delta S/m$ at tuition and fees that cover κ exactly, since (21) holds at $S \leq S^R$; no public outlay is required. With $\kappa_F = \kappa$ the real resource costs per trainee coincide, so the difference in welfare is the fiscal distortion $\phi\, s(S)\, \delta S/m$ per firm per period. The credit reform of Proposition 8 has neither. □

B. Numerical Illustration: Derivations

Senior production is $G(S) = AS^{1-\eta}/(1-\eta) = 24S^{0.25}$ with $A=6$ and $\eta = 0.75$, so $G'(S) = 6S^{-0.75}$. The remaining parameters are $\beta = 0.95$, $\beta^M = 0.90$, $\delta = 0.10$, $m = 0.10$, $\lambda = 0.30$, $\theta = 0.75$, $\omega = 1$, $c = 0.20$, $\ell = 0.8$, $a = 0.1$, $b_F = 0.1$, $w_{\min} = 0$, $b_U = 1.5$, $\bar{J} = \bar{J}^R = 5$, $\kappa = 0.25$, so that $\underline{w} = 0.6$. The maximum educational annuity payment is $(0.05/0.95)\,1.5 = 0.0789 < \omega - \ell = 0.2$, so (5) holds. Then $r = 0.145$, $r^M = 0.19$, $\beta m/r = 0.6552$, and with $x \equiv S^{-0.75}$,

$$\Delta(S) = \frac{4.5x - 1}{0.145}, \qquad v(S) = \frac{1.5x}{0.145}, \qquad v^M(S) = \frac{1.5x}{0.19}.$$

Pre-AI. $S^{\text{pre}} = m\bar{J}/\delta = 5$. $w^S(5) = 4.5 \times 5^{-0.75} = 1.346$, $\Delta(5) = 2.385$, and the market-clearing junior wage $1 - 0.095 \times 2.385 = 0.773$ exceeds the floor. Flow welfare is $G(5) - 5 - 5 = 25.89$ and $W^{\text{pre}} = 25.89/0.05 = 517.77$.

Thresholds. The floor binds below $\hat{S}$, where $0.6552(4.5x - 1) = 0.4$, giving $\hat{S} = 3.94$. Juniors are displaced in the market when $c < \hat{c} = 0.773 - 0.063 \times v^M(5) = 0.625$ and in the first best when $c < c^* = 1 - 0.6552\,[G'(5) - 1] = 0.480$; $c = 0.20$ satisfies Assumption 3.

Steady states. Each regime reduces to an equation in x. First best, from (12): $0.8 = 0.6552(6x - 1)$, so $x = 0.370$ and $S^* = 3.762$. Residency-efficient, from (23): $1.05 = 0.6552(6x - 1)$, so $S^{*,R} = 3.045$. Flexible wages with patient firms, from (13): $0.8 = 0.6552[(4.5 + 0.7 \times 1.5)x - 1]$, so $S^B = 3.391$. Flexible wages with a myopic decision-maker, from (14): $0.8 = 0.6552(4.5x - 1) + 0.063 \times 1.5x/0.19$, giving $S^{B,M} = 3.156$. Wage floor with patient firms, from (15): $0.4 = 0.688x$, so $S^F = 2.061$. Wage floor with a myopic decision-maker, from (16): $0.4 = 0.497x$, so $S^M = 1.337$. Residency, from (22): $0.6552(4.5x - 1) = 1.05 - g$, so $S^R = 2.075$ at $g = 0$, 2.564 at $g = \kappa = 0.25$, 3.762 at $g = 0.6138 = \kappa + \beta m v(S^*)$, and 3.045 at $g = 0.4263 = g^{*,R}$. The market-clearing junior

wage $1-0.6552(4.5x-1)$ is 0.48 at S^B, 0.41 at $S^{B,M}$, -0.06 at S^F, and -0.72 at S^M; the last two are hypothetical wages at floor-constrained stocks, and the flexible-wage rows are attained only where $\underline{w} \le 0.41$. Net tuition in the residency is $\kappa - c - g$: 0.05 without subsidy and -0.20, -0.56, -0.38 with the three subsidies. With $\kappa_1 = 0.05, 0.10, 0.20$ the steady-state residency stock solves $0.6552(4.5x-1) = 1.05 + \kappa_1 S$, giving 1.93, 1.81, and 1.64. Capacity: $\delta S^M/m = 1.337$ and $\delta S^R/m = 2.075$, so by (26) capacity 1 leaves $S = 1.337$, capacity 2 gives $S = 2.000$, and capacity 2.075 or more gives S^R.

Welfare and the transition. For any target $s \in (0, 5]$ let $T = \lceil \log(s/5)/\log 0.9 \rceil$ be the first date with $S_T = s$, $S_t = \max\{5(0.9)^t, s\}$, and $J_t = (S_{t+1} - 0.9S_t)/0.1$. The exact per-firm present value for a technology with extra training resource cost k is

$$W(s,k) \;=\; \sum_{t=0}^{T-1} 0.95^t \left[24S_t^{0.25} - S_t - 1 - (0.8+k)J_t\right] + \frac{0.95^T}{0.05}\left[24s^{0.25} - s - 1 - (0.8+k)s\right]. \quad (29)$$

The tail uses $J = \delta s/m = s$, and (29) includes partial hiring in the crossing period. Set $k = 0$ for direct training and $k = \kappa$ for the active residency. The resulting present values, in deviation from W^{pre}, are those in Table 1; the stock reaches S^* in period 3, S^F and S^R in period 9, and S^M in period 13. To check the equilibrium claim of Proposition 2(ii), I computed D_t and Q_t by backward recursion from the stationary tail along each path and verified the conditions of Lemma 1: $p_t - \beta^M(1-\lambda)mQ_{t+1}$ is zero to machine precision in every period with $J_t > 0$ and is nonnegative in every period with $J_t = 0$, in all four market regimes. At $S_t = 3.645$ in the flexible-wage patient regime, for instance, the premium computed from the next-period senior value is 0.2753, equal to the value of maturation $\beta(1-\lambda)mv(S^B) = 0.2753$; pricing the junior from the current stock instead would give 0.3376 and violate the condition.

Instruments, credit, and public funds. At S^*, $p(S^*) = 0.4$ and $\beta^M(1-\lambda)mv^M(S^*) = 0.1841$, so $s^* = 0.2159$ and $r^* = s^*/(\beta^M m) = 2.3987$. The market-clearing wage at $S^{B,M}$ is $w^{B,M} = 0.4100$, so credit reform requires $b_F \ge \ell - a - w^{B,M} = 0.29$ (and 0.22 for patient firms, whose clearing wage is 0.4753). For Proposition 12, at $S^R = 2.075$ the outlay per trainee beyond the verification cost is $p(S^R) - \beta^M(1-\lambda)mv^M(S^R) = 0.4 - 0.2877 = 0.1123$, or 0.233 per firm per period.

Supervision cost thresholds. Solving $S^R(\kappa)$ as κ varies and subtracting supervision costs, $W^R(\kappa) = W^M$ at $\bar{\kappa} = 0.4675$, and W^R is strictly decreasing on the active range.

The activity cutoff is $\kappa^{\text{act}} = \beta m \Delta(S^M) - \omega + c = 0.9159$; educational financing remains feasible up to this cutoff because $a + b_U - \ell + c = 1.0 > 0.9159$. As κ approaches κ^{act}, the residency's welfare converges to W^M less 12.9 of supervision costs, not to W^M. At $\kappa = 0.57$ the residency stock is 1.650 and $W^R - W^M = -3.44$. In Table 2, $\underline{w} = \ell - a - 0.1$ for $a \in \{0, 0.1, 0.2, 0.3\}$, the market stock solves (9) with the floor, and the credit needed is $\ell - a + \kappa - c$. The unsubsidized residency has $W^R = W^M$ at $\underline{w} = 0.533$ (assets 0.167) and $S^R = S^M$ at $\underline{w} = 0.488$ (assets 0.212).