Institute for New Economic Thinking
Working Papers
ineteconomics.org/workingpapers

Steering Technological Progress

Anton Korinek[†] and Joseph Stiglitz[‡]

Working Paper No. 232

May 5th, 2025

ABSTRACT

Rapid progress in new technologies such as AI has led to widespread anxiety about adverse labor market impacts. This paper asks how to guide innovative efforts so as to increase labor demand and create better-paying jobs while also evaluating the limitations of such an approach. We develop a theoretical framework to identify the properties that make an innovation desirable from the perspective of workers, including its technological complementarity to labor, the factor share of labor in producing the goods involved, and the relative income of the affected workers. Applications include robot taxation, factor-augmenting progress, and task automation. We find that steering technology becomes more desirable the less efficient social safety nets are. If technological progress devalues labor, the desirability of steering is at first increased, but beyond a critical threshold, it becomes less effective, and policy should shift toward greater redistribution. If labor's economic value diminishes in the future, progress should increasingly focus on enhancing human well-being rather than labor productivity.

https://doi.org/10.36687/inetwp232

JEL codes: E64, D63, O3

Keywords: technological progress, AI, inequality redistribution

---

[†] University of Virginia, CSH Vienna, Brookings, CEPR and NBER
[‡] Columbia University, NBER and Roosevelt Institute

We would like to thank Ken Lipartito, Pia Malaney, Joseba Martinez and participants in the CIFAR Workshop on Innovation, Equity, and the Future of Prosperity, the 2nd INET/IMF Conference on Macroeconomics in the Age of AI, the NBER Economics of Artificial Intelligence Meeting and several seminars for helpful comments. Daniel Harper provided excellent research assistance. We gratefully acknowledge financial support from the Center for Innovation, Growth and Society of the Institute for New Economic Thinking (INET-CIGS).

1 Introduction

Technological advances in recent years have led to widespread anxiety that progress will soon make an increasing number of human professions redundant. For example, Eloundou et al. (2024) predict that half of US jobs are significantly exposed to recent advances in generative AI. A substantial number of technologists go even further and predict that artificial intelligence will soon reach and then surpass human levels of general intelligence (see e.g. Kurzweil, 2005; Bostrom, 2014; Korinek and Juelfs, 2024), enabling them to perform all jobs more cheaply than the subsistence cost of human labor, and threatening to make human labor economically redundant. Such predictions are of course speculative and subject to considerable uncertainty. Nonetheless, they suggest that it may be a good idea for economists to think more carefully about how the direction of technological progress affects human well-being.

Our perspective is that technological progress does not happen by itself but is driven by human decisions on what, where, and how to innovate. It would be misplaced to view our fate as pre-determined by blind technological forces and market forces that are beyond our control, as some techno-fatalists suggest. We as a society have the power actively steer the path of technological progress in AI to address the challenges posed by our technological possibilities. Moreover, our material condition is shaped jointly by the technological innovations that we humans create and by the social and economic institutions that we collectively design and within which these innovations take place.

The central topic of this paper is thus how to steer progress in AI so as to increase demand for labor rather than displacing labor. We identify what the labor market effects of a given innovation are and how to categorize AI-based innovations according to their effects on labor demand. For this, it is necessary to pinpoint what the key conceptual properties of an innovation are that increase labor demand and therefore raise wages and employment. For example, AI-based intelligent assistants complement and augment human labor; they include navigation systems that allow unskilled workers with little geographical knowledge to take up jobs as drivers. On the other hand, technologies such as Autonomous Vehicles may predominantly substitute for workers and may lower demand for human labor. Similarly, AI agents that automate human work, such as call center agents, are likely to lower labor demand whereas agents that speed up research processes, such as in drug development, are likely to increase labor demand.

We start out by assuming that it is desirable for the economy to offer well-paying jobs to all able-bodied workers, for two complementary reasons: First, jobs offer income, and from a political economy perspective, it may be difficult in today's world to sustain the large transfers that would be required if a significant part of the work force is displaced by AI and could no longer earn a living from work. Secondly, from a psychological perspective, jobs offer not only income but also identity, pride and meaning to workers. However, we also show that the desirability of steering progress declines if the market shares the benefits of innovation more fairly or if better redistributive instruments are available.

The technical model setup that we develop builds on the approach to public economics of Atkinson and Stiglitz (1980), which solves for optimal public policy while recognizing that the private agents subject to public policy interventions also maximize their individual objective functions. However, we specifically focus on how to apply the tools of public policy to steering technological progress in AI. In doing so, we build on recent descriptions of progress with emphasis on information technologies and AI, such as Greenwald and Stiglitz (2014), Baqaee and Farhi (2018), Acemoglu and Restrepo (2018) and Korinek and Stiglitz (2018, 2019). Our main innovation over these existing works is to ask how and in which directions to actively steer technological progress to make its effects on worker as beneficial as possible.[1]

In our baseline model, we consider a framework of endogenous technological progress and assume a set of agents who differ in their exogenous factor endowments (e.g. capital and labor, or labor of different skill levels). We compare how a laissez-faire economy determines under which technologies the economy operates with what a constrained social planner would choose who values the welfare of the different agents according to defined weights. In our baseline model, we assume the planner is unable to perform transfers between the agents of the economy but can shape the economy's technology as a second-best way to affect the factor earnings of the agents, internalizing what one may call the *social pecuniary externalities* from innovation. Her optimal choice of technology depends on an innovation's complementarity to different factors and how the holdings of these factors covary with social marginal utility of the agents in the economy. In other words, the planner chooses technologies that raise the demand for factors that are owned by relatively poor agents in the economy.

In an extension of the baseline model, we consider a planner who has access to a non-linear income tax schedule but faces costs of redistribution. We find that the planner engages in more technological steering the more costly it is to redistribute. On the other hand, in the limit case of costless redistribution, steering is undesirable – the planner chooses the most efficient production technology and implements her distributive objectives via transfers. If technological progress devalues labor, it becomes at first desirable to steer against this development to enhance labor demand. Yet beyond a critical threshold, steering becomes less effective as labor is devalued, and optimal policy shifts toward greater redistribution. Throughout both phases, the devaluation of labor makes it more desirable for the planner to increase redistribution.

We provide several key applications of our theoretical framework. First, we examine robot taxation in a setting where robots can substitute for human labor. We find that a planner with sufficient welfare weight on workers will impose positive robot taxes, with the tax rate increasing in the planner's concern for workers' welfare. For more comprehensive treatments of robot taxation, see also Guerreiro et al. (2022); Costinot and Werning (2023); Thuemmel (2023); Beraja and Zorzi (2025). Second,

[1] Our analysis contributes to a long literature on endogenous and directed technological progress, going back to Ahmad (1966); Drandakis and Phelps (1966); Kennedy (1964) and Samuelson (1965). More recent works include Acemoglu (1998, 2002, 2010) and Acemoglu and Restrepo (2018).

we analyze factor-augmenting technological progress, showing that when capital and labor are gross complements (which is empirically more plausible at present), a planner concerned with workers' welfare would favor capital-augmenting innovations to raise wages. Third, we investigate the welfare-maximizing level of task automation, finding that a welfare-maximizing planner would choose to automate fewer tasks than what production efficiency would dictate when workers' welfare is heavily weighted.

We extend our analysis to economies with multiple goods and identify an important additional effect. The planner increases social welfare not only by affecting factor incomes but also by focusing technological progress on making goods cheaper that are disproportionately consumed by relatively poorer agents, thereby raising their real income. This effect would remain relevant even for agents for whom the role of labor income declines.

We further study how technological choices interact with market power. When workers have market power, profit-maximizing firms pursue innovations that erode workers' market power by making them more easily replaceable, even at the expense of production efficiency. A social planner who values worker welfare would instead employ technologies that preserve workers' market power. Similarly, when employers have monopsony power, they choose technologies that expand this power beyond what a social planner would consider optimal.

Lastly, we consider how to balance the monetary and non-monetary aspects of technological progress. We find that firms may not sufficiently account for non-monetary aspects of technological progress in both general well-being (including safety) and work. A planner would find it desirable to include such considerations in steering technological progress. Moreover, if the role of labor declines due to technological devaluation, the planner would increasingly shift attention to non-monetary considerations. In such scenarios, we show that the focus of steering would shift from labor productivity to enhancing direct utility and improving the non-monetary quality of any remaining work.

Our findings on how to steer technological progress to maximize the positive impact is relevant in five specific domains: First, many entrepreneurs in the technology sector are eager to maximize the positive impact of their developments on mankind and will find it useful to obtain better guidance on the likely impact of their developments on income distribution. If such entrepreneurs direct their innovative capabilities toward this goal, they can significantly enhance labor market outcomes for the average worker. Second our findings are useful for unions and work councils that are interested in how to steer progress to the benefit of their members. Third, a significant part of AI research is either conducted or sponsored by government. Using our findings on the labor market implications of different types of innovations, such research can actively be steered in a direction that augments human labor rather than replacing it. Fourth, our work also highlights the important role that our broader policy framework (including our tax system) plays in steering technological progress: at present, labor is the most highly-taxed factor in our economic system whereas the cost of capital has been kept low – perhaps artificially low – by long periods of expansionary monetary policy, creating strong incentives for labor-saving and capital-using innovation. One of the most natural public policy steps to steer progress in a direction that augments human labor is to reduce the burden of taxation on labor. Last but not least, our work also provides insights on how to actively provide economic incentives for innovative efforts to benefit workers.

The rest of this paper proceeds as follows. Section 2 introduces our baseline model and the main result on how to optimally steer progress when considering distributive objectives. Section 3 examines the optimal interactions between steering technology and redistributive policies. Section 4 presents three applications focused on robot taxes, factor-augmenting progress, and automation. Section 5 extends the analysis to multiple goods. Section 7 considers how technological choices interact with market power. Section 6 analyzes how to weigh the monetary and non-monetary aspects of technology. The final section concludes.

2 Model

2.1 Setup

Consider an economy in which there are $i = 1, ...I$ agents and $h = 1, ..., H$ factors of production. Each individual agent i has a utility function $u^i\left(c^i\right)$ over consumption c^i. Furthermore, each agent is born with a vector of factor endowments $\ell^i = (\ell^{i1}, ..., \ell^{iH})'$ that add up to a total factor endowment $\ell = \sum_i \ell^i$.[2]

There is also a representative firm that has access to a technology described by the production function $F(\ell; A)$ for a given vector of factor inputs ℓ and a vector of technological parameters $A = \left(A^1, \ldots, A^K\right) \in \mathbb{R}^K$, which capture in reduced form the state of technology in the economy and what investments in R&D have been made. We assume that the production function exhibits constant-returns-to-scale in the factors ℓ and that the representative firm is competitive so that it earns zero profits in equilibrium and questions of ownership are irrelevant. (The case of decreasing returns can easily be subsumed by introducing a fixed factor "ownership" that earns any excess profits.)

2.2 First Best

We start by analyzing the first-best allocation in the described economy. We consider a social planner who maximizes social welfare in the economy, given by the weighted sum of utility of individual consumers, with an exogenous set of weights $\left\{\theta^i\right\}$. W.l.o.g. we assume that the welfare weights are normalized so that $\sum_i \theta^i = 1$. This allows us to use the welfare weights to define a probability measure and an assoicated expectations operator E_i. Social welfare can then be equivalently expressed as a sum over all agents'

[2]We will investigate the additional considerations that arise with multiple goods $j = 1, ...J$ in Section 5 below.

utilities or as an expectation

$$W = \sum_i \theta^i u^i \left(c^i\right) = E_i \left[u^i \left(c^i\right)\right]$$

The planner's choices are (i) to pick the technological parameters $A = \left(A^1, \ldots, A^K\right)$ in the economy and (ii) to allocate consumption $\{c^i\}$ to the consumers in the economy – equivalent to the capacity to perform lump-sum transfers. The planner's optimization problem is

$$\max_{c^i, A} W = \sum_i \theta^i u^i \left(c^i\right) \quad \text{s.t.} \quad \sum_i c^i = F\left(\ell; A\right) \tag{1}$$

This formulation highlights that the planner's choice of technology and the consumption allocation can be performed in two separate steps. The first step is the following.

Definition 1 (Production Efficiency)**.** For given factor endowment ℓ, we denote the set of efficiency-maximizing technological parameters $A^*\left(\ell\right)$ and the associated level of output $y^*(\ell)$ so that

$$A^*\left(\ell\right) = \arg\max_A F\left(\ell; A\right) \quad \text{and} \quad y^*\left(\ell\right) = F\left(\ell; A^*\right) \tag{2}$$

For brevity of notation we will omit the argument ℓ on A^* and y^* unless required for clarity. If the technology parameters are specified such that $F\left(\ell; A\right)$ is continuously differentiable and concave in A and the maximization problem in (2) has an interior optimum, then production efficiency is described by

$$F_A\left(\ell; A\right) = 0$$

Proposition 1 (First-best allocation)**.** *For given welfare weights and factor endowments, the planner chooses the technology parameters in the economy to achieve production efficiency. She chooses the consumption allocations such that they exhaust production and satisfy the optimality conditions*

$$\theta^i u^{i\prime}\left(c^i\right) = \lambda \quad \forall i$$

Proof. The first part follows because if production efficiency was not satisfied, it would be easy to increase welfare by moving to a more efficient technology choice. The second part follows from taking the optimality conditions of the Lagrangian of the planner's maximization problem. □

The planner simply distributes resources among consumers so that their weighted marginal utilities of consumption are equated – and equal the shadow price on the economy's resource constraint.

The proposition reflects that production efficiency can be pursued independently of distributive concerns – the planner simply maximizes output and then transfers it to consumers in a desirable manner. However, there is by now a large literature explaining why the second welfare theorem is not in general a good guide for public policy. This paper can be thought of as expanding on those discussions in the context of endogenous technology.

2.3 Laissez Faire Equilibrium

In the laissez faire equilibrium, each agent i rents out her factor endowments at the prevailing rental rates $w = (w_1, ..., w_L)$ to earn a total factor income of $w \cdot \ell^i$, which she consumes. The problem of an individual consumer is thus

$$\max_{c^i} u^i\left(c^i\right) \quad \text{s.t.} \quad c^i = w\ell^i$$

where we define μ^i as the Lagrangian on the agent's budget constraint.

The representative firm rents the factors of production ℓ from the agents of the economy and picks the technology parameters A so as to maximize total profits

$$\max_{\ell, A} \Pi = F(\ell; A) - w \cdot \ell \tag{3}$$

The equilibrium in the economy consists of a set of consumption allocations $\{c^i\}$, factor allocations $\{\ell^i\}$ and technological parameters A together with rental rates w such that all agents and the representative firm satisfy their optimization problem and goods and factor markets clear, i.e. $\sum_i c^i = F(\ell; A)$ and $\sum_i \ell^i = \ell$.

Proposition 2 (Laissez-faire equilibrium)**.** *Under laissez-faire, the consumption allocations and technology parameters in the economy satisfy the optimality conditions*

$$u'\left(c^i\right) = \mu^i \quad \forall i$$

$$F_\ell(\ell; A) = w \tag{4}$$

$$F_A(\ell; A) = 0 \tag{5}$$

The laissez-faire allocation satisfies production efficiency and is Pareto efficient.

Proof. The proof follows from taking the optimality conditions of the Lagrangian of private agents' and the firm's maximization problems. The decentralized optimality conditions replicate the conditions of the first-best for appropriately chosen welfare weights $\theta^i = 1/\mu^i$ and satisfy the same constraints; therefore the allocation is Pareto efficient. □

The first optimality condition reflects that each agent allocates consumption efficiently; however, the overall distribution of wealth is determined by each agent's factor endowment, reflected in the agent's shadow value of wealth μ^i. The second condition ties factor returns to their marginal products. The last optimality condition reflects that a decentralized firm will pursue production efficiency – just like the planner in the first best.

We denote the factor shares s_ℓ earned by the different factors ℓ in the economy by

$$s_\ell(\ell; A) = \frac{F_\ell(\ell; A) \circ \ell}{F(\ell; A)}$$

where the operator $\circ$ represents the element-by-element (Hadamard) product of the two factors.

2.4 Constrained Planner

Let us now analyze a constrained planner with weights $\{\theta^i\}$ on individual utilities who is unable to perform transfers between the agents of the economy – the only way to affect the income distribution is via competitive factor returns, which depend on the choice of technology.[3] This setup serves as a benchmark to contrast to the first-best setup in section 2.2 and illustrate our basic insights in as simple a setting as possible. In Section 3, we show that the same basic insight holds if transfers are available but costly – a setting that more closely reflects the real-world situation faced by policymakers. The consumption of agent i is

$$c^i = w \cdot \ell^i = F_\ell(\ell; A) \cdot \ell^i \tag{6}$$

The constrained planner substitutes the implementability constraint (6) into her objective function and solves

$$\max_A W = \sum_i \theta^i u^i \left(F_\ell(\ell; A) \cdot \ell^i\right) \tag{7}$$

For the following proposition, we assume that the planner's optimization problem is concave in A and has an interior solution.

Proposition 3 (Constrained Optimum; No Transfers)**.** *The constrained planner chooses the technology parameters of the economy such that they satisfy*

$$\sum_i \theta^i u^{i\prime}\left(c^i\right) F_{\ell A}(\ell; A) \cdot \ell^i = 0 \tag{8}$$

Proof. The proof follows from taking the optimality conditions to the constrained planner's objective. □

Intuitively, the planner's sets the technological parameters such that she weighs the marginal effect of her technology choice on the factor earnings of agent i, captured by $F_{\ell A}(\ell; A) \cdot \ell^i$, at the welfare weight and marginal utility of each agent i. By doing so, the planner internalizes the wage effects of her technology choices, which can be viewed as *social pecuniary externalities.*

2.5 Decomposing the Effects of Technological Change

A constrained planner's choice of technology generically deviates from the benchmark of production efficiency that prevails in both the first best and the decentralized equilibrium. Let us now characterize the trade-off between efficiency and redistribution a bit further.

[3]The constrained planner's problem described below is isomorphic to the problem of a Ramsey planner who sets taxes or subsidies on the described choice variables and rebates (or raises) any associated revenue with lump sum transfers to the same set of agents from whom it was obtained.

Lemma 1 (Decomposition of (Marginal) Technological Change)**.** *For given factor inputs ℓ, the effects of a marginal technological change dA on factor returns F_ℓ can be decomposed into a* zero-sum redistribution *between factors that satisfies $\bar{F}_{\ell A} \cdot \ell = 0$ and a proportional* scale parameter *on all factor returns so that*

$$F_{\ell A} = \underbrace{\bar{F}_{\ell A}}_{\textit{redistribution}} + F_\ell \cdot \underbrace{\frac{F_A}{F}}_{\textit{scale par.}}$$

Proof. Define $\bar{F}_{\ell A} = F_{\ell A} - F_\ell \cdot F_A/F$ and observe that

$$\bar{F}_{\ell A} \cdot \ell = F_{\ell A} \cdot \ell - F_\ell \cdot \ell \frac{F_A}{F} = F_A - F_A = 0$$

Note that the last step applies Euler's theorem to each of the two terms of the sum, i.e. $F_A = F_{\ell A} \cdot \ell$ and $F = F_\ell \cdot \ell$. □

We can employ this decomposition to re-formulate the technology choice (8) of a constrained social planner in terms of the traditional equity-efficiency trade-off:

$$\underbrace{E_i \left[u^{i\prime} \left(c^i \right) \bar{F}_{\ell A} \cdot \ell^i \right]}_{\text{marg. redistributive effect}} = \underbrace{F_A E_i \left[u^{i\prime} \left(c^i \right) c^i / F \right]}_{\text{marg. efficiency effect}}$$

The left-hand side of this expression distills the redistributive effects of technology choice – reflected in the zero-sum redistribution $\bar{F}_{\ell A}$. The right-hand side captures only the efficiency effects of the technology choice – reflected in the overall change in output F_A converted into units of weighted average marginal utility.

2.6 Implementation of Constrained Optimum

Let us now consider how to implement the constrained optimum in a decentralized setting. Assume that the representative firm faces a linear tax vector τ on the choice of the technological parameters A. (W.l.o.g. we parameterize technology such that this specification of taxes is meaningful). Then the firm's profits are

$$\Pi = F(\ell; A) - w \cdot \ell - \tau \cdot A$$

and the firm's optimality condition on A becomes

$$F_A(\ell; A) = \tau \tag{9}$$

Compared to optimality condition (5), the firm deviates from production efficiency to account for the tax.

To see how to implement the constrained optimal allocation, we identify the tax τ necessary so that expression (9) replicates the constrained planner's optimality condition (8). We find

Corollary 1 (Implementation of Constrained Optimum)**.** *To decentralize the constrained social optimum, a planner imposes the following taxes on the choice of technology A,*

$$\tau = -F_{\ell A} \cdot E_i \left\{ \ell^i \left[u^{i\prime}\left(c^i\right) - E_i u^{i\prime}\left(c^i\right) \right] \right\} = -F_{\ell A} \cdot Cov_i \left(u^{i\prime}\left(c^i\right), \ell^i \right) \qquad (10)$$

Proof. We use Euler's theorem to rewrite expression (9) as

$$F_A\left(\ell; A\right) = F_{\ell A}\left(\ell; A\right) \cdot \ell = \tau$$

We then subtract equation (8) from the resulting expression to obtain

$$\tau = -\left(E_i \left[u^{i\prime}\left(c^i\right) F_{\ell A}\left(\ell; A\right) \cdot \ell^i \right] - E_i \left[u^{i\prime}\left(c^i\right) \right] F_{\ell A}\left(\ell; A\right) \cdot \ell \right)$$

Rearranging this expression results in the tax formula (10). □

Intuitively, the tax rate takes into account how much the technological parameter A benefits or hurts each factor h, captured by the cross-derivative $F_{\ell A}$, and how much those factor holdings ℓ^i covary with each agent's marginal utility. The planner will subsidize technological progress if, on average, it benefits factors that are owned by agents with comparatively high marginal utility.

2.7 Tools to Steer Technological Progress

Our results above offer a sharp analytic description of how to steer technological progress when distribution is a concern. Although we acknowledge the practical difficulties in following this approach, we view our results as a useful guidepost for what direction of technological change is desirable in at least four different settings.

First, many innovators and entrepreneurs in the technology sector are eager to maximize the positive impact of their developments on society. At present, there is a great deal of focus on how AI developers can avoid discrimination, biases, etc. – even if it comes at the expense of somewhat reducing their profit margins (see e.g. Dubber et al., 2020). However, the impact of technological progress on labor markets and income distribution is all too often an afterthought for innovators. Publicly-spirited innovators will find it useful to be reminded of and obtain better guidance on the likely impact of their inventions on workers. If the world's most creative innovators put their minds to it, they can play an important positive role in guiding progress in a direction that is beneficial for the average worker. Furthermore, innovators are perhaps also best-suited to predict the potential implications of their innovations and make better-informed decisions on what innovations to pursue to further the interests of workers.

Second, unions and works councils may have a say in which types of investments and innovations to pursue in their companies, and they may also be well-suited in judging the effects of specific innovations on workers. If they have the right to participate in the decision-making process, they will steer technological progress in a direction that is positive for their members. This is the precise opposite of the efforts of some corporations to make their workers as replaceable as possible in order to reduce workers' bargaining power. Moreover, it may also counteract the tendency of management to automate workers because machines are seen as easier to manage and maintain, even if such a move comes at the expense of production efficiency.[4]

Third, a significant part of AI research is either conducted or sponsored by government. Although this type of research is funded by the tax dollars of all workers, the government typically pays little attention to how the resulting innovations affect the livelihoods of all workers. A natural public policy is to evaluate the likely labor market effects of innovations when determining what type of research the government should pursue or fund.

Fourth, the tax formulas that we derived above would be the most direct instruments to guide technological progress in a desirable direction. However, more generally, our tax system plays an outsized role in affecting the direction of technological progress – whether intentionally or unintentionally: at present, labor is the most highly-taxed factor in our economic system, creating strong incentives for labor-saving innovation (see e.g. Acemoglu et al., 2020). One of the most natural public policy measures to steer progress in a direction that augments human labor is to reduce the burden of taxation on labor or to even subsidize human labor.

3 Costly Redistribution

We now extend our analysis to incorporate the possibility of redistribution. We introduce non-linear income taxation with administrative costs. Our generalization allows us to examine the trade-off between efficiency and redistribution in a more realistic setting while maintaining analytical tractability.

3.1 Setup with Non-Linear Income Taxation

We build on the single-good baseline model, where agent i has utility $u^i(c^i)$ and factor endowments ℓ^i. The pre-tax income of agent i is $y^i = F_\ell(\ell; A) \cdot \ell^i$. After applying the non-linear tax $T(y^i)$, agent i's consumption becomes

$$c^i = y^i - T(y^i)$$

We introduce costs associated with implementing the tax system, which can be interpreted as either administration costs or as a reduced-form representation of potential distortions generated by taxation. (As we showed in the analysis of the first-best above, concerns about distribution would be moot if the planner has a tool to costlessly redistribute among agents. Since factor endowments are exogenous in our setting, non-linear income taxation would be able to achieve the first-best in many cases.) We

[4]In a first-best world, such over-automation would be penalized by a reduction in managerial compensation. However, in a world with agency frictions, managers may use their discretion to advantage their well-being at the expense of their workers and shareholder.

assume that taxation costs arise at the level of the individual taxpayer and are given by $\gamma(T(y^i))$where $\gamma'(T) \geq 0$ and $\gamma''(T) \geq 0$, capturing the increasing marginal cost of taxation, with strict inequalities for some relevant values of T. For example, $\gamma(T)$ could be zero or small for positive transfers to agents $T < 0$, but rising and convex for taxes raised from agents $T > 0$. The government budget constraint then captures that net revenues need to cover administration costs,

$$\sum_i T(y^i) = \sum_i \gamma(T(y^i)) \tag{11}$$

3.2 Constrained Planner

The constrained planner maximizes

$$\max_{A,\{T(\cdot)\}} W = \sum_i \theta^i u^i(F_\ell(\ell; A) \cdot \ell^i - T(F_\ell(\ell; A) \cdot \ell^i))$$

subject to (11). The Lagrangian for this problem is

$$L = \sum_i \theta^i u^i(y^i - T(y^i)) + \lambda \sum_i \left[T(y^i) - \gamma(T(y^i))\right]$$

Taking the first-order conditions with respect to $T(y^i)$ and A yields

$$\theta^i u^{i\prime}(c^i) = \lambda(1 - \gamma'(T(y^i)))$$

$$\sum_i \theta^i u^{i\prime}(c^i) \cdot (1 - T'(y^i)) \cdot F_{\ell A}(\ell; A) \cdot \ell^i + \lambda \sum_i [T'(y^i) \cdot F_{\ell A}(\ell; A) \cdot \ell^i \cdot (1 - \gamma'(T(y^i)))] = 0$$

The first condition $\theta^i u^{i\prime}(c^i) = \lambda(1 - \gamma'(T(y^i)))$ implicitly pins down the optimal tax function. Note that when $\gamma'(T) = 0$ (no administrative costs), we return to the first-best condition with $\theta^i u^{i\prime}(c^i) = \lambda$ for all agents, achieving perfect redistribution according to the planner's welfare weights.

Using the first expression to substitute out $\lambda(1 - \gamma'(T(y^i)))$ in the second expression and simplifying, we obtain

$$\sum_i \theta^i u^{i\prime}(c^i) \cdot F_{\ell A}(\ell; A) \cdot \ell^i = 0$$

which is identical to the optimality condition (8) in the baseline model. This demonstrates that the basic insight of our baseline model – that the planner chooses technology according to its marginal-utility-weighted impact on the factor earnings of different agents, continues to hold.

The Efficiency-Redistribution Trade-off Alternatively, by substituting the optimality condition $\theta^i u^{i\prime}(c^i) = \lambda(1 - \gamma'(T(y^i)))$ again and applying Euler's theorem, we can also express the optimality condition for the technology choice as

$$F_A(\ell; A) = \sum_i \gamma'(T(y^i)) \cdot F_{\ell A}(\ell; A) \cdot \ell^i \tag{12}$$

This formulation reveals the central trade-off between production efficiency and redistribution:

1. When $\gamma'(T) = 0$ (no costs of redistribution), we have $F_A(\ell; A) = 0$, corresponding to production efficiency as in the first-best. Redistributive goals are achieved via transfers.

2. When $\gamma'(T) > 0$, the planner generally deviates from production efficiency. The deviation depends on the sum of marginal costs $\gamma'(T(y^i))$ of taxing agent i weighted by how much the technology choice A affects agent i's factor income. The deviation from production efficiency is higher the greater the cost of taxes and transfers.

This highlights that the optimal technology balances the costs of inefficient production against the costs of taxation. As the marginal cost of taxation increases, the planner relies more heavily on technology choice as a redistributive tool.

Our extension demonstrates that the fundamental insight of our baseline model – that a constrained planner deviates from the efficiency-maximizing technology choice when redistribution is difficult to perform – continues to hold, while modeling a more realistic constraint on redistribution. Unlike the original model in Section 2, where the planner cannot perform any transfers between agents, here the planner can implement transfers but faces costs for doing so. In the limiting cases of $\gamma'(T) \to 0$, we approach the first-best solution with efficient production and optimal redistribution. As $\gamma'(T) \to \infty$ for transfers, we obtain the constrained optimum when transfers are unavailable, replicating the result of Proposition 3.

3.3 Steering Versus Redistribution When Factors Are Devalued

As progress in AI continues to accelerate, it raises the prospect of some factors being progressively devalued. For instance, specific types of labor or skills, and one day, perhaps labor as a whole, may be increasingly substituted by AI, reducing their marginal product. This possibility raises an important question in the context of our model: how should the social planner adjust the balance between steering technological progress and direct redistribution when the value of certain factors, particularly the labor of disadvantaged agents, declines?

We extend the model with costly redistribution from the previous subsection by introducing an exogenous factor devaluation parameter $\delta \in [0, 1]$ that affects the marginal product of a certain factor h, where $\delta = 1$ represents full value and lower δ indicate greater devaluation, with $\delta = 0$ representing full devaluation. We denote the production function as a function of δ by $F(\ell; A, \delta)$.[5] The pre-tax income of agent i continues to be $y^i = F_\ell(\ell; A, \delta) \cdot \ell^i$, and the constrained planner's problem as well as the optimality conditions of the planner remain structurally unchanged.

To analyze how factor devaluation affects the optimal balance between steering and redistribution, we examine how changes in δ influence the optimal tax schedule and technology choice. W.l.o.g. let us focus on a technology parameter A that positively affects the marginal product of factor h, i.e., $F_{\ell^h A}(\ell; A, \delta) > 0$, and let us assume that factor h is predominantly owned by disadvantaged agents, amounting to $\text{Cov}_i(u^{i\prime}(c^i), \ell^{ih}) > 0$. Under this constellation, "increased steering" toward factor h corresponds to choosing higher values of A, while "decreased steering" corresponds to choosing values of A, closer to the production-efficient level, which is typically lower than the constrained social optimum when factor h is owned by disadvantaged agents. We find:

Proposition 4 (Factor Devaluation and Steering Vs. Redistribution)**.** *Consider an economy where factor h is disproportionately owned by disadvantaged agents with higher weighted marginal utility and is subject to devaluation parameterized by δ.*

(i) Optimal redistribution monotonically increases as factor devaluation increases (δ decreases) so $\partial T^*(y^i)/\partial\delta > 0$ for disadvantaged agents and conversely for advantaged agents who do not suffer from factor devaluation.

(ii) The optimal steering toward factor h exhibits an inverted U-shaped pattern around a critical threshold $\delta^* \in (0, 1)$. For mild devaluations of factor h, i.e., $d\delta < 0$ when $\delta > \delta^*$, make steering more desirable ($dA^*/d\delta < 0$). For more severe devaluations of labor, i.e., $d\delta < 0$ when $\delta < \delta^*$, steering becomes less desirable ($dA^*/d\delta < 0$). *The transition threshold δ^* is characterized by $dS_h(\delta)/d\delta\big|_{\delta=\delta^*} = 0$* where

$$S_h(\delta) = Cov_i\left[\gamma'(T^i(\delta)), \ell^{ih}\right]$$

represents the 'tax-distortion-weighted factor holdings' of factor h.

(iii) In the limit as $\delta \to 0$ (complete devaluation), the optimal technology choice approaches production efficiency with respect to factor h:

$$\lim_{\delta \to 0} F_{\ell^h A}(\ell; A^*, \delta) \cdot \ell^h = 0 \tag{13}$$

Proof. (i) To prove the monotonic increase in redistribution, we first observe that as δ decreases, the pre-tax income of disadvantaged agents (y^i) decreases relative to that of advantaged agents, increasing inequality. From the optimality condition $\theta^i u^{i\prime}(c^i) = \lambda(1 - \gamma'(T(y^i)))$, as y^i decreases, $u^{i\prime}(c^i)$ increases. To maintain optimality, $\gamma'(T(y^i))$ must decrease, implying a decrease in $T(y^i)$ given the convexity of γ. Conversely, for advantaged agents j, $T(y^j)$ must increase to satisfy the government budget constraint (11).

[5] To provide a specific example, if the effective supply of factor h is given by $\hat{\ell}^h = \delta\ell^h$ and enters the production function as a gross substitute, then the marginal product $F^h_\ell = \delta \cdot \partial F/\partial\ell^h$ declines towards zero as $\delta \to 0$.

(ii) From the optimality condition (12), the deviation from production efficiency depends on the covariance between the tax distortion $\gamma'(T(y^i))$ and how technology affects each agent's income via $F_{\ell A}(\ell; A, \delta) \cdot \ell^i$. For factor h, this relationship can be characterized by $S_h(\delta) = \text{Cov}_i[\gamma'(T^i(\delta)), \ell^{ih}]$. As δ decreases from 1, two competing effects arise. First, the marginal utility of disadvantaged agents increases, raising the value of steering toward factor h. Second, the effectiveness of steering declines as factor h is devalued, reducing the income gains from favorable technology choices.

Initially, the first effect dominates, leading to increased steering as δ decreases (i.e., $dA^*/d\delta < 0$ for $\delta > \delta^*$). At the critical threshold δ^*, these effects balance exactly, satisfying $dS_h(\delta)/d\delta\big|_{\delta=\delta^*} = 0$. For $\delta < \delta^*$, the second effect dominates, and steering decreases as devaluation increases further (i.e., $dA^*/d\delta > 0$).

To formally derive the transition point, we totally differentiate the analogue of equation (12) expanded by δ with respect to δ, noting that at the optimal technology choice A^*, the left-hand side equals the right-hand side. The sign of $dA^*/d\delta$ is determined by $dS_h(\delta)/d\delta$, which changes from negative to positive at δ^*.

(iii) In the limit as $\delta \to 0$, factor h becomes economically irrelevant as$\lim_{\delta \to 0} F_{\ell^h}(\ell; A, \delta) = 0$. Consequently, steering technology toward factor h yields no benefit, so the optimal choice approaches production efficiency with respect to this factor. □

The proposition delivers several important insights about optimal policy responses to the devaluation of production factors. First, it confirms that as factors owned by disadvantaged agents lose value, the desirability of redistributive policies for the planner monotonically increases. The planner accepts progressively higher costs of redistribution as inequality widens to maximize welfare.

More interestingly, the proposition reveals a non-monotonic relationship between factor devaluation and the optimal degree of technological steering. This relationship follows an inverted U-shaped pattern: as devaluation begins, the optimal policy initially involves increased steering toward the devalued factor (higher A^* under our assumptions), but after reaching a critical threshold, the optimal policy shifts away from steering and toward greater reliance on redistribution. (If we had instead assumed that technology substitutes for factor h (i.e., $F_{\ell^h A} < 0$), then the directions would be reversed, but the fundamental non-monotonic pattern would remain.) The key insight is the existence of a transition point where the balance between steering and redistribution shifts.

The intuition behind this pattern is the following. When factors begin to lose value, steering technology to complement these factors becomes more socially valuable because of the higher marginal utility of the affected agents. However, as devaluation becomes more severe, the effectiveness of steering diminishes—even the most favorable technology choices cannot significantly improve incomes derived from nearly worthless factors. Eventually, the efficiency costs of deviating from production efficiency outweigh the diminishing distributional benefits.

The "tax-distortion-weighted factor holdings" $S_h(\delta) = \text{Cov}_i[\gamma'(T^i(\delta)), \ell^{ih}]$ provides a precise way to identify this transition point. This measure captures how the marginal cost of taxation covaries with holdings of the devalued factor, measuring the distributional value of steering technology toward that factor. When $dS_h(\delta)/d\delta = 0$, the optimal intensity of steering reaches its maximum.

In the context of AI-driven labor market transformations, these results suggest a nuanced policy approach. Initially, as certain labor types begin to lose value, policymakers should focus on steering technological progress to complement these workers—for instance, developing AI that enhances rather than replaces their productivity. However, if devaluation progresses beyond a critical point, policy emphasis should shift toward more direct redistributive measures, reflecting the diminishing returns to technological steering when factors approach economic obsolescence.

This framework helps resolve an apparent tension in the literature on technological displacement. Some argue for steering technology to benefit disadvantaged workers, while others emphasize strengthening redistribution mechanisms. Our analysis suggests both approaches are optimal but at different stages of factor devaluation. The key insight is that the balance between steering and redistribution should evolve dynamically as the value of the affected factors changes, rather than remaining fixed.

In future economic scenarios where the value of human labor may be significantly impaired by advanced AI, understanding this transition will be crucial for designing effective policy responses that maintain welfare while acknowledging the practical constraints on both technological steering and redistribution.

4 Applications

4.1 Robot Taxation

The first and simplest application of our baseline model captures a key concern in the Age of AI: what happens when technological progress creates machines that can directly replace workers? Mapping this application into our baseline model, we assume an economy where machines can substitute for human labor but policymakers have an instrument to discourage their use so their technological choice is $A = \tau \geq 0$, i.e., they pick the level of a "robot tax."

We consider a production function where robots R perfectly substitute for labor L,

$$F(K, L + R) = K^\alpha (L + R)^{1-\alpha} \tag{14}$$

with $\alpha \in (0, 1)$. Robots can be employed at a fixed contemporaneous cost $\eta > 0$ per unit.

We assume that there is a unit mass of capitalists, each owning ℓ^K unit of capital, and a unit mass of workers, each owning ℓ^L units of labor. The total factor endowments in the economy are thus $K = \ell^K$ and $L = \ell^L$.

First-Best and Decentralized Equilibrium In the first-best allocation, a social planner would maximize total output net of robot costs

$$\max_{R\geq 0} K^{\alpha}(L+R)^{1-\alpha} - \eta R \tag{15}$$

If the constraint $R \geq 0$ is slack, then the planner's first-order condition holds with equality, $F_L = \eta$. We define the cost threshold above which zero robots are optimal by $\eta^* = F_L(K, L) = (1-\alpha)K^{\alpha}L^{-\alpha}$. For $\eta < \eta^*$, a positive number of robots $R > 0$ is employed to satisfy $F_L = \eta$ or, equivalently,

$$R(\eta) = \left(\frac{(1-\alpha)K^{\alpha}}{\eta}\right)^{1/\alpha} - L > 0 \tag{16}$$

In that case, total output (gross of robot costs) is $Y(\eta) = F(K, L+R(\eta)) = [(1-\alpha)/\eta]^{(1-\alpha)/\alpha}K$. Otherwise, for $\eta \geq \eta^*$,no robots are employed, $R(\eta) = 0$. The first-best allocation satisfies production efficiency. The planner allocates the output to capitalists and workers according to their relative welfare weights $\{\theta^K, \theta^L\}$ so that $\theta^K u^{K\prime}(c^K) = \theta^L u^{L\prime}(c^L)$.

In the decentralized equilibrium, workers and capitalists supply their factor endowment. A representative firm hires K, L, and $R(\eta)$ robots and pays the marginal products F_K and F_L. The availability of robots places a ceiling on workers' wages $w \leq \eta$. The decentralized equilibrium allocation coincides with the first-best for welfare weights

$$\hat{\theta} = \frac{\theta^L}{\theta^K} = \frac{u^{K\prime}\left(\alpha Y(\eta)\right)}{u^{L\prime}\left(\eta L\right)}$$

Constrained Social Planner Now consider a constrained planner who cannot directly redistribute income but can impose a tax $\tau \geq 0$ on robots, with the tax proceeds rebated lump-sum to workers.[6] With the tax, robots cost $\eta + \tau$ per unit, inducing private firms to employ $R(\eta + \tau)$ units.

The planner's objective function is

$$\max_{\tau \geq 0} \theta^K u^K\left(w_K\left(\ell;\tau\right)K\right) + \theta^L u^L\left(w_L\left(\ell;\tau\right)L + \tau R\right) \tag{17}$$

We solve the planner's problem in three steps. First, we observe that for $\theta^L/\theta^K < \hat{\theta}$, the planner finds it optimal to impose zero taxes, replicating the decentralized equilibrium. Second, if $\theta^L = 1$ so the planner places all her weight on workers, the optimum maximizes workers' consumption, consisting of wages plus robot tax revenue, and solves

$$\max_{\tau}(\eta+\tau)L + \tau R(\eta+\tau)$$

[6]This is a natural assumption since the objective of the tax intervention is to make workers better off. Although the distributive benefit for workers declines the less of a share of the tax revenue they receive, the basic insight holds for other forms of distributing the revenue raised. In particular, even if all the revenue was distributed to capitalists, taxing robots in our framework would still help workers and be desirable for a planner with a sufficient welfare weight on workers.

with optimality condition

$$L + R(\eta + \tau) + \tau R'(\eta + \tau) = 0$$

At $\tau \approx 0$, we observe that the left-hand side is positive, but as τ rises, the third term is negative and becomes, at first, progressively more so. If the optimum tax is an interior solution, then it is reached when the marginal benefit of higher wages and revenue from robot taxes (captured by the first two terms) is precisely offset by the marginal cost of discouraging robots with higher taxes. Conversely, if the tax rises to the level where $\eta + \tau = \eta^*$,then it is optimal for the policymaker to impose a prohibitive tax on robots, w.l.o.g., $\tau^* = \eta^* - \eta$. We summarize our findings as follows:

Proposition 5 (Robot Taxation)**.** *For a sufficiently low weight on workers, $\theta^L/\theta^K \leq \hat{\theta}$, the constrained planner imposes zero robot taxes $\tau = 0$. For $\theta^L/\theta^K > \hat{\theta}$, the planner's optimal robot tax τ is positive and strictly increasing in workers' relative weight θ^L/θ^K up to τ^*, which is the optimal tax rate in the limit of $\theta^L = 1$.*

Proof. See discussion above. □

Intuitively, as the planner places more weight on workers' welfare, she increases the robot tax to raise wages, even though this sacrifices some production efficiency. The tax creates a transfer from robot users to workers through both higher wages and additional tax revenue. The application underlines the key theme of our paper: that it is optimal to deviate from production efficiency when distributional concerns are important and direct transfers are limited.

The effects of imposing taxes on robots are closely related to quantity regulations on how much robot use is admissible.

4.2 Factor-Augmenting Progress

Our second application considers factor-augmenting technological progress. In its pure form, factor-augmenting progress implies that the same amount of output can be produced using less input of the augmented factor. Examples of labor-augmenting innovation include intelligent assistants that enable a given worker to perform her duties more efficiently, or more efficient techniques of managing workers so that a given amount of labor can effectively provide more labor services. An example of capital-augmenting innovation is the progress in semiconductor technology that is captured by Moore's Law, whereby a given quantity of silicon chips can perform ever more computation.

Formally, we assume that the function $a_j(A)$ of the technology parameter A determines how much technology augments factor ℓ_j so that $a_j(A)\,\ell_j$ effective units of it enter production. We collect these functions in a vector $a(A) = (a_j(A))_j$ and denote by $a(A) \circ \ell$ the vector of effective factor inputs, where $\circ$ is the Hadamard (element-by-element) product. The production function is then $F(\ell; A) = F(a(A) \circ \ell)$.

Two-Factor Production Function Let us consider a general production function with two factors, capital K and labor L, with factor-augmenting technology entering as

$$y = F(a_K(A)K, a_L(A)L)$$

where F is twice continuously differentiable, increasing in both arguments and concave. This captures a wide range of production technologies, including CES. Assume the parameter A reflects a trade-off between augmenting capital and labor, with $a'_K(A) > 0$ and $a'_L(A) < 0$.

For concreteness, consider an economy in which there are only two types of individuals, capitalists K and workers L, who are endowed with one unit of capital and labor, respectively. In a slight abuse of notation, we label them by $i = K, L$ and denote their endowments $\ell^K = (1, 0)'$ and $\ell^L = (0, 1)'$ so the economy's total factor endowment is $\ell = (1, 1)$.

Decentralized Equilibrium A representative firm maximizes profits (3). The firm's first-order conditions determine the competitive factor rents for $h = K, L$,

$$w_K = F_1(\cdot)a_K(A) \quad \text{and} \quad w_L = F_2(\cdot)a_L(A)$$

Varying the parameter A traces out the economy's innovation possibilities frontier in the space (w_L, w_K). We assume this frontier is concave, reflecting decreasing returns to directing technological progress toward either factor. The relative ratio of rents is given by

$$R(A) = \frac{w_K}{w_L} = \frac{F_1(\cdot)a_K(A)}{F_2(\cdot)a_L(A)}$$

We now provide a local characterization of the relative rent function by employing the elasticity of substitution, which measures how the ratio of effective factor inputs changes locally in response to changes in relative marginal products,[7]

$$\sigma = \sigma(\text{K,L}) = \frac{d\ln(a_K(A)K/a_L(A)L)}{d\ln(F_2/F_1)}$$

Log-differentiating $R(A)$ with respect to A gives

$$\frac{d\ln R}{dA} = \frac{d\ln(a_K/a_L)}{dA} + \frac{d\ln(F_1/F_2)}{dA}.$$

By the definition of the elasticity of substitution,

$$\frac{d\ln(F_1/F_2)}{dA} = -\frac{1}{\sigma}\frac{d\ln(a_K/a_L)}{dA}.$$

[7]The first equality emphasizes that the elasticity is a local property of the production function that depends on the factor inputs (K, L). For the special case of CES production functions, the elastictiy is constant across all factor inputs pairs.

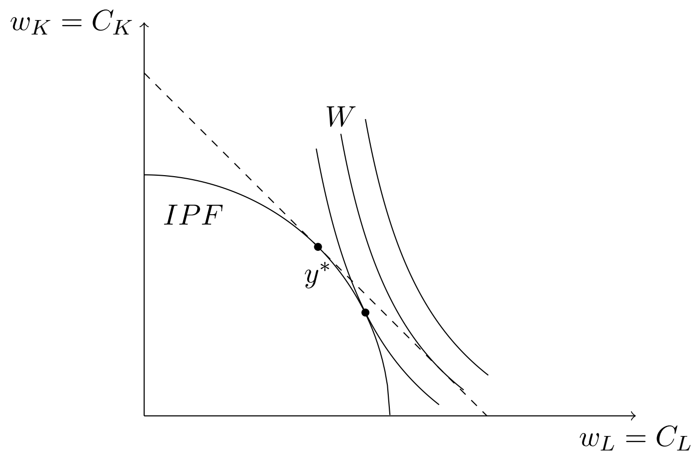

Figure 1: Innovation possibilities frontier and welfare isoquants

Hence

$$\frac{d\ln R}{dA} = \left(1 - \frac{1}{\sigma}\right)\frac{d\ln(a_K/a_L)}{dA} = \frac{\sigma - 1}{\sigma}\,\frac{d\ln(a_K/a_L)}{dA}. \tag{18}$$

Since $a'_K(A) > 0$ and $a'_L(A) < 0$, we have $d\ln(a_K/a_L)/dA > 0$. Consequently, when factors are gross complements ($\sigma < 1$), augmenting capital by increasing A decreases the ratio of capital returns to wages, benefiting workers. When factors are gross substitutes ($\sigma > 1$), the opposite occurs, and increasing A raises the ratio of capital returns to wages, hurting workers. For a unitary elasticity ($\sigma = 1$), changes in A leave relative factor prices unchanged.

Constrained Planner A constrained planner who optimizes the social welfare function (7) solves

$$\max_A \theta^K u^K(w_K(A)) + \theta^L u^L(w_L(A)) \tag{19}$$

The first-order condition for this problem balances the marginal welfare effects of changing technology for capitalists and workers, weighted by their respective welfare weights

$$\theta^K u^{K\prime}(w_K) \cdot \frac{dw_K}{dA} + \theta^L u^{L\prime}(w_L) \cdot \frac{dw_L}{dA} = 0 \tag{20}$$

The planner's problem is illustrated in Figure 1, in which the economy's innovation possibilities frontier is illustrated by a concave locus. The dashed line with slope −1 represents the efficiency-maximizing choice of technology. By contrast, the planner chooses the point on the frontier at which it forms a tangent to her welfare isoquants (convex curves in the figure), guaranteeing the highest level of welfare possible.

We now show our main result for this example analytically:

Proposition 6 (Factor-Augmenting Progress). *If factors are gross complements ($\sigma < 1$), then the constrained planner's optimal choice of A is a strictly increasing function of the relative weight on workers versus capitalists θ^L/θ^K. If factors are gross substitutes ($\sigma > 1$), the opposite results apply.*

Proof. Using the relative ratio of rents function R, we find $w_K = R\,w_L$ or, differentiated w.r.t. A, $dw_K/dA = dR/dA \cdot w_L + dw_L/dA \cdot R$. This allows us to express the planner's optimality condition (20) as

$$\frac{dw_L}{dA} = -\frac{\theta^K u^{K\prime}\, w_L}{\theta^K u^{K\prime}\, R + \theta^L u^{L\prime}} \cdot \frac{dR}{dA},$$

so $\text{sgn}(dw_L/dA) = -\text{sgn}(dR/dA)$.

Denoting the left-hand side of (20) by $G(\theta^L, \theta^K, A)$ and applying the implicit function theorem, we find

$$\frac{dA}{d(\theta^L/\theta^K)} = \frac{\partial G/\partial(\theta^L/\theta^K)}{\partial G/\partial A} = -\frac{u^{L\prime}(w_L)\cdot \frac{dw_L}{dA}}{\text{SOC}}$$

where SOC is the second-order condition, which satisfies $SOC < 0$ for interior solutions. Therefore $\text{sgn}(dA/d(\theta^L/\theta^K)) = \text{sgn}(dw_L/dA) = -\text{sgn}(dR/dA)$, which is given by expression (18), proving the proposition. □

The left-hand side of the planner's optimality condition is decreasing in θ^L/θ^K. For the empirically more plausible case where factors are gross complements ($\sigma < 1$), workers benefit from higher A (more capital augmentation), while capitalists are hurt. The right-hand side is decreasing in A for this case.

Intuitively, the more weight the planner places on the welfare of workers versus capitalists, the more she wants to gear technological progress in a direction that raises wages relative to capital rents and generates a redistribution from capitalists to workers. If the factors are gross complements, this can be done by augmenting capital relatively more than labor; if they are gross substitutes, it requires augmenting labor relative to capital. The planner's willingness to deviate from the first-best solution depends both on the efficiency cost and the curvature of her welfare isoquants, i.e., her desire to redistribute.

Examples of factor-augmenting technologies One example of a labor-augmenting technology is intelligent assistants, which are frequently cited by AI developers as holding promise for improving the productivity of workers. These are AI-powered devices that assist workers and increase their productivity by complementing their cognitive capabilities. A specific example of such assistants are Augmented-Reality devices that help to upskill lesser-skilled workers by providing them with instructions on how to perform cognitively intensive jobs. Such devices can assist factory workers perform complicated workflows that would otherwise require significant training. Another application is AI systems that provide call center workers with additional information about the callers, e.g. by analyzing the emotional content of voices. Even navigation systems can be interpreted as intelligent assistants that augment human drivers and allow them to navigate more efficiently and to navigate routes in areas that they are not familiar with.[8]

Platforms that match labor services can be interpreted as another example of labor-augmenting innovations. A number of high-tech corporations specialize in matching demand and supply for labor in the economy. An important example are ride-sharing platforms, which match demand and supply for drivers.[9] Taxi drivers used to spend significant amounts of time looking for jobs, and the matching efficiency of these platforms has enabled them to deliver more driving services in the same amount of time—while also devaluing their human capital. Other examples include MTurk, which matches demand and supply for tasks that human workers can perform digitally, and Etsy, which matches demand and supply for artisan goods.

Whether labor-augmenting technological progress ultimately benefits workers depends on the elasticity of substitution/demand for labor, as highlighted in our proposition. If the elasticity is less than unity, then productivity increases raise the effective supply of labor by more than they raise demand for it, resulting in lower returns to labor. This is what seems to have happened e.g. in the ride-sharing market. If the elasticity is above unity, returns to labor rise.

4.3 Automation of Tasks

Our next application considers a setup that centers on the question of task automation in a framework that is inspired by Zeira (1998) and Acemoglu and Restrepo (2019). We assume that final output is produced using a unit mass $j \in [0, 1]$ of intermediate goods or tasks according to the production function $\log y = \int_0^1 \log y(j)\, dj$, where each individual task is performed using either capital K or labor L. Specifically, we assume the variable $A \in [0, 1]$ captures what fraction of the tasks is automated so they are performed using capital and $y(j) = K(j)$ for $j \leq A$. The remaining fraction $(1 - A)$ reflects all the tasks that are not automated and are performed using labor so $y(j) = L(j)$ for $j > A$. As in our previous application, we assume that capital is owned exclusively by capitalists and labor by workers. We denote the fractions of the economy's factor endowments of capital and labor by $K/L = \alpha/(1 - \alpha)$.

Within the set of automated and non-automated tasks, it is optimal to allocate capital and labor symmetrically. For given K and L, this implies that $y(j) = {}^{K}/_{A}$ for $j \leq A$ and $y(j) = {}^{L}/_{1-A}$ for $j > A$. The aggregate production of the economy can then

[8] We also note an important potential downside of intelligent assistants: they may actually lower the skill levels of workers because they make them dependent on the assistants, they may thus turn human workers that used to think for themselves more and more into "robots" that mechanically follow the instructions given by the assistant.

[9] There are justified concerns about the job conditions of workers at ride-sharing companies; these concerns are in addition to the effects of these platforms on labor demand and could be addressed separately by appropriate regulation.

be expressed as

$$F(K, L; A) = \left(\frac{K}{A}\right)^A \left(\frac{L}{1-A}\right)^{1-A} \tag{21}$$

Lemma 2. *Production efficiency implies that a fraction $A = \alpha$ of tasks is automated. This is what is replicated by the laissez faire equilibrium.*

Proof. Maximizing the log production function (21) delivers the optimality condition $\frac{A}{\alpha} = \frac{1-A}{1-\alpha}$ which is satisfied for $A = \alpha$. Proposition 2 implies that the same holds in the laissez faire equilibrium. □

By contrast, production efficiency will generally no longer hold if we are concerned with the distributive implications of automation and if direct transfers are not available. To see this, we solve the problem of a second-best planner who maximizes the welfare function (19).

Proposition 7 (Task Automation)**.** *The constrained planner chooses a degree of automation A strictly between the welfare weight θ^K on capitalists and the fraction α of the factor endowment that is capital. An increase in the welfare weight on workers reduces the optimal degree of automation.*

Proof. See appendix. □

Intuitively, the fraction A of automated tasks also represents the share of output that is earned by capitalists in a decentralized setting. A planner who places greater weight on workers will reduce automation, which increases the fraction of tasks available for workers and raises their share of output. However, deviating from production efficiency reduces the total amount of output. At the optimum, the planner weighs off the desired redistribution with the associated loss in production efficiency.

Examples of Task Automation Task automation occurs when a machine acts as a perfect substitute for a task in a productive process that was previously performed by labor. A tangible example is an assembly line that consists of a series of steps performed by humans, and a machine is introduced to perform one of them. Choosing the level of automation A in our analytic framework then corresponds to deciding how many of the steps are automated and how many are performed by humans. In the given example, tasks are perfect complements in the sense that each task along the assembly line is required in fixed proportion to produce the output.

More generally, tasks may also be combined in a more elastic fashion. When machines substitute for tasks performed with labor and simultaneously increase the productivity at which the task is performed, then this can also be thought of as task-augmenting progress akin to the factor augmentation in the previous section. If tasks are gross complements (with elasticity less than one), then augmenting the automated task will benefit other tasks performed by labor. For example, if doctors produce health services both by diagnosing and by providing advice in a complementary fashion, then automating diagnosis and making it more efficient may actually increase their returns from providing advice. If the elasticity is below one, the opposite result applies. In the production function (21) above, tasks are combined in Cobb-Douglas fashion, i.e. with unitary elasticity, so no such effects occur.

Acemoglu and Restrepo (2018) observe that the fraction A of automated tasks can be affected in two ways: by changing the degree of automation of existing tasks or by creating new tasks. Accordingly, choosing a lower degree of automation than the decentralized equilibrium can be achieved not only by holding back the automation of existing tasks performed by human labor but also by inventing new human-focused tasks that are performed by labor. However, for many new inventions, it is unclear if they will enter the production function in the particular form specified in (21) or if they will give rise to more fundamental changes to the productive structure of the economy.

5 Multiple Goods Economy

Having examined specific applications of our framework to robot taxation, factor-augmenting progress, and task automation in the context of a single-good economy, we now extend our analysis to incorporate multiple goods and sectors. This allows us to identify an additional mechanism through which technological progress affects welfare: by impacting the relative prices of different consumption goods.

Consider an economy with $i = 1, ..., I$ agents, $j = 1, ..., J$ goods, and $h = 1, ..., H$ factors of production. Each agent i has utility function $u^i(c^i)$ over consumption bundle $c^i = (c^{i1}, ..., c^{iJ})'$ and factor endowments $\ell^i = (\ell^{i1}, ..., \ell^{iH})'$ that add up to a total factor endowment $\ell = \sum_i \ell^i$.

Each sector j has a production function $y_j = F^j(\ell^j; A^j)$, where ℓ^j is the vector of factors used in sector j and A^j are sector-specific technological parameters, collected in the vector $A = (A^1, ..., A^J)$. We assume that each production function exhibits constant-returns-to-scale in the factor inputs. The resource constraints for factors and goods require

$$\sum_j \ell^j = \ell \qquad \text{and} \qquad \sum_i c^{ij} = F^j(\ell^j; A^j) \quad \forall j \tag{22}$$

First-Best Allocation A social planner with welfare weights $\{\theta^i\}$ (normalized so that $\sum_i \theta^i = 1$) maximizes

$$\max_{\{c^i\},\{\ell^j\},A} W = \sum_i \theta^i u^i(c^i)$$

subject to the resource constraints (22). The Lagrangian for this problem is

$$\mathcal{L} = \sum_i \theta^i u^i(c^i) + \sum_j \lambda_j \left[F^j(\ell^j; A^j) - \sum_i c^{ij} \right] + \mu \cdot \left[\ell - \sum_j \ell^j \right]$$

where λ_j is the shadow price of good j and μ is the vector of shadow prices for factors. Taking first-order conditions, we can find that the first-best allocation satisfies production efficiency within each sector and an optimal allocation of factors across sectors so weighted marginal factor products are equated,

$$F^j_{A^j}(\ell^j; A^j) = 0 \quad \forall j$$
$$\lambda_j F^j_{\ell^h}(\ell^j; A^j) = \mu_h = \lambda_k F^k_{\ell^h}(\ell^k; A^k) \quad \forall j, h, k$$

Moreover, the planner optimally distributes consumption so that weighted marginal utilities are equated,

$$\theta^i u^i_{c^{ij}}(c^i) = \lambda_j \quad \forall i, j$$

Decentralized Equilibrium We assign the price vector $p = (p_1, ..., p_J)$ for goods and continue to use $w = (w_1, ..., w_H)$ for factors. Each consumer i and each firm j solve

$$\max_{c^i} u^i(c^i) \quad \text{s.t.} \quad p \cdot c^i = w \cdot \ell^i \tag{23}$$
$$\max_{\ell^j, A^j} \Pi^j = p_j F^j(\ell^j; A^j) - w \cdot \ell^j \tag{24}$$

The decentralized equilibrium allocation satisfies consumer optimality,

$$u^i_{c^{ij}}(c^i) = \mu^i p_j \quad \forall i, j$$

where μ^i is the Lagrange multiplier on consumer i's budget constraint, implying:

$$\frac{u^i_{c^{ij}}(c^i)}{u^i_{c^{ik}}(c^i)} = \frac{p_j}{p_k} \quad \forall i, j, k$$

Firm optimality implies

$$p_j F^j_{\ell^h}(\ell^j; A^j) = w_h \quad \forall j, h$$
$$F^j_{A^j}(\ell^j; A^j) = 0 \quad \forall j$$

and market clearing is given by (22).

The laissez-faire allocation satisfies production efficiency but generally fails to maximize social welfare for defined welfare weights due to an undesirable distribution of consumption.

Constrained Social Planner's Problem A constrained planner who cannot perform transfers between agents can influence welfare only through the choice of technology A. The planner internalizes that consumers maximize utility (23) given prices and factor endowments, firms maximize profits given prices and technology, modifying (24) so that A^j is no longer a choice variable, and markets clear (22).

Formally, we represent the equilibrium consumption and factor allocations and prices by $c^i(A)$, $\ell^j(A)$, $p(A)$, and $w(A)$ as a function of the planner's choice of technology A. The planner's problem is then

$$\max_A W = \sum_i \theta^i u^i(c^i(A))$$

subject to the stated constraints.

Proposition 8 (Constrained optimum with multiple goods). *The constrained planner chooses technology parameters to satisfy*

$$\sum_i \theta^i u^i_c(c^i) \cdot \left[\frac{\partial (w(A) \cdot \ell^i)}{\partial A^{jk}} - \frac{\partial (p(A) \cdot c^i(A))}{\partial A^{jk}} \right] = 0 \quad \forall j, k$$

where $u^i_c(c^i)$ is the marginal utility of income for agent i. By the envelope theorem and budget constraint, this simplifies to

$$\sum_i \theta^i u^i_c(c^i) \cdot \left[\sum_h \ell^{ih} \frac{\partial w_h}{\partial A^{jk}} - \sum_m c^{im} \frac{\partial p_m}{\partial A^{jk}} \right] = 0 \quad \forall j, k \tag{25}$$

Proof. Taking the derivative of the welfare function with respect to A^{jk},

$$\frac{\partial W}{\partial A^{jk}} = \sum_i \theta^i \sum_m u^i_{c^{im}}(c^i) \frac{\partial c^{im}(A)}{\partial A^{jk}}$$

From consumer i's optimization, we know $u^i_{c^{im}}(c^i) = \mu^i p_m$. Substituting

$$\frac{\partial W}{\partial A^{jk}} = \sum_i \theta^i \mu^i \sum_m p_m \frac{\partial c^{im}(A)}{\partial A^{jk}}$$

From the budget constraint $p \cdot c^i = w \cdot \ell^i$, we get

$$\sum_m p_m \frac{\partial c^{im}(A)}{\partial A^{jk}} + \sum_m c^{im} \frac{\partial p_m}{\partial A^{jk}} = \sum_h \ell^{ih} \frac{\partial w_h}{\partial A^{jk}} + \sum_h w_h \frac{\partial \ell^{ih}}{\partial A^{jk}}$$

Since factor endowments are fixed, $\frac{\partial \ell^{ih}}{\partial A^{jk}} = 0$. Substituting

$$\frac{\partial W}{\partial A^{jk}} = \sum_i \theta^i \mu^i \left[\sum_h \ell^{ih} \frac{\partial w_h}{\partial A^{jk}} - \sum_m c^{im} \frac{\partial p_m}{\partial A^{jk}} \right]$$

Setting $u^i_c(c^i) = \theta^i \mu^i$ and equating to zero yields the result. □

Our findings reflect that the social pecuniary externalities considered by the planner now also extend to relative goods prices. The optimality condition (22) balances the effects of technological change on factor earnings and its effects on consumption expenditure, weighted by each agent's marginal contribution to social welfare. Note that if the second part of the brackets is zero (as is the case when there is a single good or when consumers have identical homothetic preferences), the condition reduces to the optimality condition (8) in our baseline model.

Proposition 9 (Implementation with Multiple Goods)**.** *To decentralize the constrained social optimum, the constrained planner imposes the tax rates*

$$\tau^{jk} = -Cov_i\left(u_c^i(c^i), \sum_h \ell^{ih}\frac{\partial w_h}{\partial A^{jk}} - \sum_m c^{im}\frac{\partial p_m}{\partial A^{jk}}\right)$$

where Cov_i denotes the covariance over agents weighted by θ^i.

Proof. The proof follows from combining the optimality conditions of the planner and firms under technology taxation. □

As indicated by the formula, the planner subsidizes technological investments that either increase the factor earnings or reduce the consumption cost of agents with high social marginal utility. Observe that the latter element is relevant no matter what the source of consumers' incomes. In a future in which labor markets lose in importance in providing income, steering innovation would still be desirable to make the necessities consumed by lower-income consumers more affordable.

6 Steering Progress under Imperfect Competition

We now examine the effects of market power on incentives for steering technological progress. The following examples go beyond our baseline model, in which we assumed a competitive environment to examine the resulting considerations. The general insight is that actors who seek greater market power reduce economic efficiency and move the economy inside the Pareto frontier to achieve an allocation at which they are better off.

6.1 Specialization and Labor's Market Power

The following application captures firms' tradeoff of how specialized of a production process they choose versus how much market power their hirees will enjoy. In general, highly specialized production processes may yield significant productivity gains but also imply that the firms rely on specialized and/or highly skilled labor, which enjoys greater market power than undifferentiated unskilled labor.

Consider an economy with a single final good and a unit mass $i \in [0,1]$ of agents who are consumer-workers. Each agent i derives CES utility from consumption $u(c) =$

$c^{1-\sigma}/\left(1-\sigma\right)$ and elastically supplies specialized labor of type i subject to a disutility $d\left(\ell\right) = \ell^{1+\psi}/\left(1+\psi\right)$ with Frisch elasticity ψ. We assume that $\sigma < 1$ so that the substitution effects from wage changes dominate any income effects.

There is a representative firm in the economy that hires labor $h \in [0,1]$ for a unit mass of tasks and combines them according to the production function

$$y = A\left(\eta\right)\int_0^1 \left(\ell^h\right)^{1-\alpha} dh$$

The parameter $\eta \in [0,1]$ reflects the degree of specialization of labor that the firm chooses for the production process and simultaneously drives how much market power workers enjoy. We assume that $A\left(\eta\right)$ is strictly increasing and concave in η, i.e. specialization makes production more efficient but at decreasing speed. Moreover, we assume that the range is $A\left(\eta\right) \in \left[\underline{A}, \overline{A}\right]$, and that the function satisfies the two Inada conditions $\lim_{\eta\to 0} A'\left(\eta\right) = \infty$ and $A'\left(1\right) = 0$. However, the downside for the representative firm is that more specialization gives more monopoly power to workers. At $\eta = 0$, productivity is at its lowest level $\underline{A}$ and labor is completely unspecialized, so all types of labor are perfect substitutes and individual workers do not have any market power. Conversely, at $\eta = 1$, productivity is at its highest level $\overline{A}$, but each type of labor i is specific for a particular task $h = i$ so each agent i enjoys significant monopoly power. Intermediate levels of specialization imply that there is some limited substitutability between different types of labor. For example, at $\eta = 1/2$, each task $h \in [0,1]$ can be accomplished by precisely two agents $i, j \in [0,1]$, and the two supply labor in Cournot fashion so, in a symmetric equilibrium, they internalize that each supplies a fraction $\eta = 1/2$ of the labor within each of their sectors of employment.

The optimization problem of consumer-worker i is thus given by

$$\max_{c^i,\ell^i} u^i\left(c^i\right) - d\left(\ell^i\right) \quad \text{s.t.} \quad c^i = w\left(\eta\ell^i + \left(1-\eta\right)\ell^{\setminus i}\right)\cdot \ell^i$$

where $\ell^{\setminus i}$ denotes the supply of labor by all agents other than agent i in the agent's sectors of employment. The agent's optimality condition is then

$$w\left(1-\eta\epsilon_{w,\ell}\right) = \frac{d'\left(\ell^i\right)}{u'\left(c^i\right)} = \left(\ell^i\right)^{\psi}\left(c^i\right)^{\sigma}$$

where $\epsilon_{w,\ell} = -\frac{dw}{d\ell}\cdot\frac{\ell^h}{w}$ is the inverse demand elasticity for labor of firms, which reflects by what percentage wages need to go down for firms to demand one percent more labor. This defines an inverse demand relationship $w\left(\ell^i;\eta\right)$ with the derivative w.r.t. specialization

$$\frac{\partial w}{\partial \eta} = \frac{\epsilon_{w,\ell}\cdot\left(\ell^i\right)^{\psi}\left(c^i\right)^{\sigma}}{\left(1-\eta\epsilon_{w,\ell}\right)^2} > 0 \tag{26}$$

The representative firm hires labor and picks the technology parameters $A\left(\eta\right)$ to maximize total profits

$$\max_{\ell,\eta,A(\eta)} \Pi = A(\eta) \int_0^1 \left(\ell^h\right)^{1-\alpha} dh - \int_0^1 w^h(\eta)\, \ell^h dh$$

Observe that the representative firm is small and has no effect on the overall labor demand faced by each agent i. Therefore it acts competitively in labor markets in the sense that the wage does not depend on the quantity of labor that it hires. However, the firm internalizes that the degree of monopoly power enjoyed by the labor that it is hiring is endogenous and depends on its choice of specialization η, as captured by a wage function $w^h(\eta)$ for each variety h. The firm's optimality condition of labor for a given degree of specialization η is

$$(1-\alpha) A(\eta) \left(\ell^h\right)^{-\alpha} = w^h(\eta)$$

which implies an inverse demand elasticity $\epsilon_{w,\ell} = -\frac{dw}{d\ell} \cdot \frac{\ell}{w} = \alpha$. In a symmetric equilibrium, the optimal choice of specialization can be rewritten as

$$A'(\eta)\, \ell^{1-\alpha} = w'(\eta)\, \ell$$

The left-hand side captures the marginal efficiency gain from specialization and is strictly decreasing in η from infinity to zero The right-hand side reflects the marginal rise in labor costs associated with greater specialization, where $w'(\eta)$ is given by equation (26) and is increasing in η. The condition therefore yields a unique solution for the optimum level of specialization.

Proposition 10 (Steering Progress and Employee Market Power)**.** *The greater the weight θ^L placed on workers, the more specialized the production technology that the planner will employ.*

Proof. The firm's optimality condition is

$$A'(\eta)\, \ell^{1-\alpha} = w'(\eta)\, \ell$$

The marginal productivity gain, $A'(\eta)\, \ell^{1-\alpha}$, is strictly decreasing in η, while the marginal wage cost due to increased labor market power, $w'(\eta)\, \ell$, is increasing. Thus, there is a unique optimal η. When the planner places a greater weight θ^L on workers, the social cost of granting market power is reduced, shifting the optimum to a higher η. Hence, a larger θ^L leads to more specialized production. □

Discussion A tangible example of this result is that firms have incentives to de-skill jobs so that workers are more replaceable and have less bargaining power. If a given worker is the only one who can do a certain job, she can extract significant surplus; if anyone can do the job, then workers are perfect substitutes and are paid competitive wages. For example, the introduction of highly standardized production processes, say the conveyor belt or work procedures in the fast food industry, can be interpreted along these lines. More generally, this result reflects that there may be a broad set of innovations that do not increase productivity but that make jobs more undifferentiated and unskilled so as to reduce workers' bargaining power.

6.2 Monopsony Power in Factor Markets

Next we consider a setup in which firms have monopsony power in factor markets. We assume that each factor h is supplied by a single type of consumer-worker with CES consumption utility $u(c) = c^{1-\sigma}/(1-\sigma)$ who elastically supplies type h labor subject to a disutility cost $d^h(\ell) = \ell^{1+\psi}/(1+\psi)$ with Frisch elasticity ψ. The resulting optimization problem is $\max_{\ell^h} u\left(w^h \ell^h\right) - d^h\left(\ell^h\right)$, with optimality condition $w^h = d'\left(\ell^h\right)/u'\left(c^h\right)$. This gives rise to an inverse labor supply function

$$w^h\left(\ell^h\right) = \left(\ell^h\right)^{\frac{\psi+\sigma}{1-\sigma}}$$

with elasticity $\epsilon^h_{w,\ell} = \frac{\psi+\sigma}{1-\sigma}$. We assume that $\sigma < 1$ so that the substitution effects from wage changes dominate any income effects.

Assume a set of oligopsonistic firms, for which the extent of market power in factor markets is described by a vector $\alpha^m = \left(\alpha^{m1}, \ldots, \alpha^{mH}\right)$, where each α^{mh} captures what fraction of the demand for factor h derives from the firm. The optimization problem of firm m is

$$\max_{A,\ell^m} F\left(\ell^m; A\right) - w\left(\ell^m + L\right) \cdot \ell^m$$

where we denote by L the labor demand from all other firms and observe that for each factor h, we find $\alpha^{mh} = \frac{\ell^{mh}}{\ell^{mh}+L^h}$. The firm's optimality condition for labor demand equates marginal product to marginal revenue,

$$F_\ell(\cdot) = w + w'(\cdot)\,\ell^m = w\left(1 + \alpha^m \epsilon_{w,\ell}\right)$$

Similarly, the firm's optimal choice of technology is given by

$$F_A(\cdot) = 0$$

Specific examples of the ways in which firms increase their monopsony power in labor markets are (i) to put no-compete clauses in employment contracts, which prevent other employers in the same sector to compete for them and (ii) to provide training to workers in ways that are not easily portable to other firms.

7 Non-Monetary Benefits of Steering Technology

Our analysis thus far has examined steering technological progress primarily through the lens of monetary outcomes, focusing on how innovations affect labor demand, wages, and income distribution. However, technological innovations, particularly those involving AI, have impacts that extend far beyond their effects on factor incomes, including direct effects on well-being, autonomy, dignity, and social cohesion. Ensuring that advanced AI systems align with broader societal values requires understanding these non-monetary dimensions of welfare (Korinek and Balwit, 2024; Acemoglu, 2024). Technological innovations can generate substantial externalities that are not mediated through market transactions and therefore not captured in traditional economic analyses.

The importance of non-monetary considerations grows as technological progress accelerates. Advanced AI systems increasingly make decisions that affect human autonomy, shape social interactions, and determine access to opportunities. Bengio et al. (2025) and Jones (2024) emphasize that these externalities could even reach existential proportions if sufficiently powerful AI systems are developed. Moreover, if technological progress eventually reduces the marginal product of human labor, non-monetary dimensions of welfare will become relatively more important in evaluating technological innovations.

In this section, we expand our model to incorporate three critical dimensions of non-monetary benefits: direct effects of technology on utility beyond consumption possibilities; impacts on the quality of work experiences; and how these considerations interact with the potential devaluation of labor, changing the focus of optimal steering as the labor market's role in distributing income may diminish.

7.1 Direct Utility Effects of Technology

To incorporate the direct utility effects of technology, we now consider a utility function that depends directly on the technology parameters, denoted by $u^i(c^i; A)$. The direct impact of technology on utility can be either positive or negative. For example, privacy-protecting AI may enhance utility directly, while surveillance technologies may reduce it.

First-Best Allocation The welfare-maximizing allocation with welfare weights $\{\theta^i\}$ solves the analog of optimization problem (1). Unlike in our baseline model, the first-best solution is no longer characterized by production efficiency ($F_A(\ell; A) = 0$). Instead, the optimality condition for technology choice can be expressed as

$$F_A(\ell; A) = -\frac{1}{\lambda} \sum_i \theta^i u^i_A(c^i; A) = -\sum_i \frac{u^i_A(c^i; A)}{u^i_c(c^i; A)}$$

Intuitively, the productivity benefits of technology precisely balance its direct utility effects across all individuals. When technology directly enhances utility ($u^i_A > 0$), the first-best would push technology beyond the production-efficient level. Conversely, if technology has negative direct utility effects, the optimum would restrict it below the production-efficient level. Note that the welfare weights cancel out in the final expression because at the optimum, the planner has already allocated consumption to equalize welfare-weighted marginal utilities of consumption across all agents.

Laissez-Faire Equilibrium In the market economy, consumers maximize $u^i(c^i; A)$ subject to their budget constraint $c^i = w \cdot \ell^i$. However, the technology parameters A

are chosen by firms who maximize profits (3) and chose production efficiency as entailed by $F_A(\ell; A) = 0$. Firms do not internalize the direct utility effects of their technology choices. This is a classic externality that creates an inefficiency in competitive markets.

Constrained Social Planner Solving an analog of problem (7), a constrained planner who cannot perform transfers not only internalizes how technology affects agents' consumption levels but also how it affects their utility. The resulting first-order condition for technology choice is

$$\sum_i \theta^i \left[u^i_c(c^i; A) \cdot F_{\ell A}(\ell; A) \cdot \ell^i + u^i_A(c^i; A)\right] = 0$$

This condition balances three considerations: (i), the efficiency effects of technology choice, (ii) the distributive effects through factor returns and (iii) the direct utility effects of technology.

Efficient Market Outcomes and Implementation Under certain conditions, markets can efficiently mediate technology choices even when technology directly affects utility. This occurs when the utility effects of technology are fully captured in consumers' willingness to pay for products that embody the technology. As observed by Lancaster (1966), if technology determines product attributes and these attributes are fully observable to consumers and can be prized, then competitive markets will direct innovation toward characteristics that consumers value most highly.

However, this efficiency requires several stringent conditions. First, consumers must have perfect information about how technology affects product quality and their utility. Second, the utility effects must impact only the purchasing consumer with no externalities beyond the transaction. Third, markets must be competitive so firms are responsive to consumer willingness to pay. Fourth, there can be no public good aspects to the technology where benefits accrue to non-purchasers.

In practice, these conditions frequently fail. Many technologies create externalities affecting non-purchasers, such as social media platforms whose design choices affect even non-users through societal impacts. Many technologies have effects that consumers cannot fully evaluate in advance, such as AI algorithms with complex and opaque decision-making processes. Many technological characteristics create public goods or bads that markets systematically do not internalize.

When markets fail to achieve efficient outcomes, policy intervention becomes necessary. To implement the constrained optimum, the planner imposes taxes or subsidies on technology choices,

$$\tau = -F_{\ell A} \cdot \text{Cov}_i \left[u^i_c(c^i; A), \ell^i\right] - \frac{E_i[\theta^i u^i_A(c^i; A)]}{E_i[u^i_c(c^i; A)]}$$

The first term, as in our baseline model, addresses distributional concerns. The second term, new to this extension, corrects for the direct utility externalities of technology.

This highlights that technological steering should account for both distributional effects and direct utility impacts. Some practical implications:

Privacy-enhancing technologies may be underproduced in a market economy since firms don't fully internalize their utility benefits to users. Conversely, addictive technologies may be overproduced in markets, as firms don't internalize the negative utility impacts of addiction. Similarly, social media algorithms optimized for engagement may harm mental health, representing a negative externality that warrants correction.

Technologies with large direct utility benefits for disadvantaged populations deserve particular encouragement. Accessible design features in digital interfaces create significant utility gains for users with disabilities, but these benefits may not be fully captured in market transactions. The value of these benefits extends beyond monetary compensation, affecting dignity, autonomy, and quality of life.

7.2 Technology, Work Amenities, and Labor Supply

We now further extend our model to incorporate how technology affects the disutility of work and the labor supply decision. We assume that the utility function takes the form

$$U^i = u^i(c^i; A) - d^i(\ell^i; A)$$

where $d^i(\ell^i; A)$ represents the disutility of supplying labor, which depends both on the quantity of labor supplied ℓ^i and the technology parameters A. Labor supply is now an endogenous choice variable.

First-Best Allocation The social planner's problem becomes

$$\max_{c^i,\ell^i,A} W = \sum_i \theta^i [u^i(c^i; A) - d^i(\ell^i; A)] \quad \text{s.t.} \quad \sum_i c^i = F\left(\sum_i \ell^i; A\right)$$

Setting up the corresponding Lagrangian with shadow price λ, the first-order condition on A can be written as

$$F_A(\ell; A) = -\frac{1}{\lambda}\sum_i \theta^i [u^i_A(c^i; A) - d^i_A(\ell^i; A)] = -\sum_i \frac{u^i_A(c^i; A) - d^i_A(\ell^i; A)]}{u^i_c(c^i; A)}$$

In short, the planner finds it optimal to deviate from production efficiency based on both the direct utility effects and the work disutility effects of technology, expressed in terms of marginal utility.

Laissez-Faire Equilibrium Consumers choose consumption and labor supply to maximize

$$\max_{c^i,\ell^i} [u^i(c^i; A) - d^i(\ell^i; A)] \quad \text{s.t.} \quad c^i = w \cdot \ell^i,$$

arriving at the optimality condition

$$w \cdot u_c^i(c^i; A) = d_\ell^i(\ell^i; A)$$

Firms maximize profits as before, neglecting direct externalities and implementing production efficiency.

The market equilibrium now incorporates the effect of technology on labor supply decisions, but firms still fail to internalize both the direct utility effects and the work disutility effects of their technology choices.

Work Amenities and Market Efficiency Under certain conditions, markets can correctly value work amenities despite the externality. Specifically, if work amenities are fully reflected in the marginal disutility of labor $d_\ell^i(\cdot)$ and are observable and priced in a competitive market, then the equilibrium wage will adjust to reflect the value of these amenities, $w = d_\ell^i(\ell^i; A)/u_c^i(c^i; A)$.

However, these conditions rarely hold in practice. Markets frequently fail to efficiently provide work amenities under real-world conditions. Information asymmetries imply that workers cannot fully observe or evaluate all aspects of technology that affect their work experience before accepting employment. Limited mobility due to switching costs prevents workers from freely choosing employers based on technology-related amenities, weakening competitive pressures to provide optimal amenity levels. When employers have monopsony power in labor markets, they systematically underprovide amenities relative to the efficient level. Additionally, technologies that affect collaborative work environments create externalities across workers that individual labor supply decisions don't capture, leading to inefficient market outcomes.

Constrained Social Planner The constrained planner recognizes that labor supply responds endogenously to wages and technology, $\ell^i = \ell^i(w(A), A)$, and maximizes

$$\max_A W = \sum_i \theta^i [u^i(w(A) \cdot \ell^i(w(A), A); A) - d^i(\ell^i(w(A), A); A)]$$

The resulting optimality condition balances productive efficiency, distributive effects, direct utility effects, and work disutility. To implement the constrained optimum, the planner can impose a technology tax or subsidy,

$$\tau = -\frac{E_i[\theta^i(u_A^i - d_A^i)]}{E_i[u_c^i]} - F_{\ell A} \cdot \text{Cov}_i[\theta^i u_c^i(\cdot), \ell^i)]$$

The first term corrects for both direct utility and work disutility externalities, while the second addresses efficiency and distributional concerns. Technologies that improve overall utility or work experience (positive u_A^i or negative d_A^i) would receive subsidies, while those that deteriorate utility and working conditions face taxes.

This extension highlights important non-monetary dimensions of technological progress. The meaning and dignity of work represent critical non-monetary dimensions of technological choice. Technologies that enhance workers' sense of purpose and dignity create value beyond wages. For example, AI systems that handle routine tasks while leaving creative and meaningful components to humans may be underprovided by markets. Similarly, technologies that preserve worker autonomy generate important non-monetary benefits, while monitoring technologies that create psychological stress may be overprovided in markets when firms fail to internalize these costs.

Opportunities to develop greater skills constitute another dimension of technological impact. Technologies that facilitate learning and skill development create non-monetary benefits through human capital accumulation and increased worker engagement. The psychological benefits of mastery and growth contribute significantly to worker well-being.

Health and safety implications of technology also deserve consideration beyond their productivity impacts. Technologies that reduce physical strain, repetitive stress injuries, or exposure to hazardous conditions generate substantial utility beyond their impact on measured productivity and wages. The quality of work experience fundamentally shapes individuals' overall well-being.

7.3 Labor Devaluation and the Changing Focus of Steering

We now incorporate our analysis of labor devaluation from Section 3.3 into this expanded framework of non-monetary effects. While our earlier analysis examined the optimal balance between steering technology and redistribution when labor is devalued, we now examine how this devaluation affects the relative importance of monetary and non-monetary dimensions of technological progress.

Recall that we introduced a factor devaluation parameter $\delta \in [0, 1]$, where lower values indicate greater devaluation of labor. As labor becomes increasingly devalued (δ decreases), there are two implications for non-monetary benefits. First, the relative importance of non-monetary benefits in overall welfare increases as monetary benefits from labor decline. Second, the nature of optimal steering shifts away from focusing on labor productivity and toward focusing on direct utility effects and the quality of non-work time.

Formally, we can express the constrained planner's problem with the parameter δ as

$$\max_A W = \sum_i \theta^i [u^i(F_\ell(\ell; A, \delta) \cdot \ell^i + T^i; A) - d^i(\ell^i(\delta); A)]$$

Importantly, labor supply $\ell^i(\delta)$ is now endogenous and depends on both wages (which are affected by δ) and the technology parameters A. To make this explicit, we can write $\ell^i(\delta) = \ell^i(w(\delta), A)$ where $w(\delta) = F_\ell(\ell; A, \delta)$. We also include a transfer T^i, which can be modeled as in Section 3.3, to ensure positive income for all agents. Each agent chooses labor supply to maximize utility

$$\max_{\ell^i} u^i(w(\delta)\ell^i; A) - d^i(\ell^i; A)$$

with associated first-order condition

$$w(\delta)u^i_c(w(\delta)\ell^i + T^i; A) = d^i_\ell(\ell^i; A)$$

This captures that as δ decreases and wages decline, individuals will continue supplying labor only if the marginal disutility of labor $d^i_\ell(\ell^i; A)$ decreases sufficiently, or if they derive sufficiently high marginal utility from consumption.

The constrained planner's first-order condition with respect to the technology parameter A is

$$\sum_i \theta^i \left[u^i_c(c^i; A) \cdot \frac{\partial(F_\ell(\ell; A, \delta) \cdot \ell^i)}{\partial A} + u^i_A(c^i; A) - d^i_A(\ell^i; A) \right] = 0$$

To analyze how optimal steering changes with δ, we need to understand how the relative importance of each term is affected. The first term captures the monetary benefits of technology through factor earnings, while the second and third terms capture non-monetary benefits through direct utility effects and work disutility, respectively.

We can establish the following proposition:

Proposition 11 (Non-Monetary Benefits and Labor Devaluation)**.** *Consider an economy where labor is subject to devaluation parameterized by δ and agents derive both monetary and non-monetary benefits from technology.*

(i) As δ decreases, the optimal technology choice $A^*(\delta)$ places progressively less weight on raising the productivity of labor and more weight on direct utility effects and work amenities.

(ii) For positive marginal disutility of labor at zero ($d^i_\ell(0; A) > 0$ for all i), there exists a critical threshold $\delta^*_i > 0$ for each agent below which they supply zero labor. At $\delta = 0$ (complete devaluation), $\ell^i = 0$ for all i and the optimal technology choice A^* satisfies

$$\sum_i \theta^i [u^i_A(0; A)] = 0$$

(iii) If there exist agents with positive marginal utility of labor at zero ($d^i_\ell(0; A) < 0$ for some i, meaning intrinsic enjoyment of work), these agents continue to supply positive labor even as $\delta \to 0$. In this case, as $\delta \to 0$, the optimal technology choice A^* satisfies

$$\sum_i \theta^i [u^i_A(0; A) - d^i_A(\ell^i_0; A)] = 0$$

where $\ell^i_0 > 0$ is the labor supply chosen by agents with $d^i_\ell(0; A) < 0$ when $\delta = 0$.

Proof. (i) Let $G(\delta) = \frac{\sum_i \theta^i u_c^i(c^i;A)\cdot\partial(F_\ell(\ell;A,\delta)\cdot\ell^i)/\partial A}{\sum_i \theta^i[u_A^i(c^i;A)-d_A^i(\ell^i;A)]}$ represent the ratio of marginal monetary benefits to marginal non-monetary benefits at the optimal technology choice. Taking the derivative with respect to δ, we find $G'(\delta) > 0$ for $\delta < \delta^*$, where δ^* is the threshold identified in Proposition 4. This means that as δ decreases, the relative weight on non-monetary benefits increases.

(ii) For agents with $d_\ell^i(0;A) \geq 0$, the first-order condition $w(\delta)u_c^i(w(\delta)\ell^i;A) = d_\ell^i(\ell^i;A)$ implies that as δ decreases and $w(\delta)$ approaches zero, there exists a threshold δ_i^* below which the marginal utility of consumption cannot justify positive labor supply. When $\delta = 0$, if all agents have $d_\ell^i(0;A) > 0$, they all choose $\ell^i = 0$. The planner's first-order condition then simplifies to $\sum_i \theta^i[u_A^i(0;A)] = 0$, as the labor supply and disutility terms drop out.

(iii) For agents with $d_\ell^i(0;A) < 0$, the first-order condition can be satisfied with positive labor supply even when $w(\delta) = 0$. These agents work for the intrinsic enjoyment of labor itself. As $\delta \to 0$, the planner's first-order condition includes both direct utility effects and work disutility effects for these agents, yielding $\sum_i \theta^i[u_A^i(0;A) - d_A^i(\ell_0^i;A)] = 0$. □

This proposition reveals several important insights about how optimal technology steering changes when labor is economically devalued but non-monetary benefits remain relevant. First, as labor's economic value diminishes, the relative importance of direct utility effects and work amenities in determining optimal technology choice increases. Second, the proposition highlights an important distinction between two types of labor: that performed largely for economic rewards and that performed partly for intrinsic enjoyment.

When labor is completely devalued economically, individuals with standard preferences (positive disutility of labor) will cease working entirely. For these individuals, technology steering focuses exclusively on enhancing direct utility. However, for individuals who derive intrinsic enjoyment from work (negative marginal disutility at zero labor), some amount of labor will be supplied even when it generates no monetary value. For these individuals, technology steering must still consider both direct utility effects and the technology's impact on the work experience itself. This aligns with the observations of Korinek and Juelfs (2024), who note that as AI technology advances and potentially devalues human labor, society may need to reorient toward technological development that enhances human flourishing directly rather than through labor markets.

The implication is that a planner would increasingly focus on technologies that enhance utility outside the labor market, such as those improving health, leisure activities, social connections, and overall life satisfaction, as these factors become increasingly central to human welfare. Moreover, for anyone who derives meaning and purpose from work activities, she would make the remaining work more intrinsically rewarding, even if it is no longer economically valuable.

In summary, incorporating non-monetary benefits into our framework for steering technological progress reveals additional dimensions along which technological choices create externalities that markets fail to address. As technology increasingly affects well-being directly and potentially devalues traditional labor input, steering must account for these broader impacts.

In the near term, this calls for encouraging technologies with positive externalities on direct utility while discouraging those with negative externalities on welfare or work quality. Developing institutions to better assess and value non-monetary impacts becomes increasingly important as these dimensions gain relative significance in overall welfare.

In the longer term, if labor becomes increasingly devalued, the focus of steering shifts toward ensuring that technology enhances human flourishing broadly, rather than narrowly focusing on labor market outcomes. This suggests a complementary relationship between the two described approaches: steering technology toward labor-friendly directions in the short term while preparing for potentially deeper changes in how well-being is produced and distributed. The incorporation of non-monetary benefits into our framework helps bridge these approaches by highlighting dimensions of welfare that remain important regardless of labor's centrality to income distribution.

8 Conclusions

Throughout economic history, technological change has displaced specific categories of workers, and the ongoing rapid progress in AI may expand the range of occupations experiencing such pressures. Our social safety nets are only partially effective in countering the resulting distributional challenges. Faced with these developments, this paper has analyzed how to actively steer technological progress to have more desirable distributive effects.

We have developed a formal framework that analyzes a constrained social planner for whom redistribution is limited or costly. Our analysis shows that while production efficiency is optimal when lump-sum transfers are available, a constrained social planner without perfect redistributive instruments will generally deviate from efficiency to achieve better distributive outcomes. The extent of this deviation is proportional to the costliness of redistribution—an insight that provides theoretical grounding for when and how much to steer technological development.

Our applications illustrate these principles: robot taxation becomes increasingly desirable if a planner's welfare weight on workers rises; capital-augmenting innovations benefit workers when capital and labor are gross complements; and a planner would reduce task automation from the most efficient level to benefit workers. These findings offer concrete guidance for evaluating technological innovations.

When extending our analysis to multiple goods, we find that it is also desirable to steer technological progress toward making goods cheaper that are disproportionately consumed by relatively poorer agents, thereby raising their real income. This effect remains relevant even for agents whose labor income declines, highlighting an additional channel through which technological steering impacts welfare.

Our evaluation of imperfect competition analyzes additional considerations. When workers have market power, profit-maximizing firms pursue innovations that erode this power even at the expense of production efficiency—a tendency a social planner would counteract. Similarly, when employers have monopsony power, they choose technologies that enhance this power beyond what is socially optimal. These market imperfections create additional rationales for steering technological progress.

Beyond monetary considerations, we have shown that non-monetary aspects of technological choice are critically important. Firms may systematically undervalue how technology affects the meaning, satisfaction, and fulfillment derived from work because these factors are not efficiently priced by markets. As labor's monetary value potentially diminishes with continued technological progress, the relative importance of these non-monetary factors increases, suggesting a progressive shift in focus from labor productivity to well-being more broadly.

The tools for steering progress that we have outlined can help with both short-term and long-term challenges. In the short term, we should actively steer technological progress toward labor-using innovations to increase demand for human labor. This buys valuable time and improves the welfare of workers while our economy adjusts. Looking forward, however, we must also recognize that the nature of our challenge may evolve. If technological progress continues along its current trajectory, and in particular, if artificial general intelligence is reached, the marginal product of labor for an increasing fraction of workers could decline, and may reach levels that are insufficient to support a decent standard of living (Korinek and Juelfs, 2024). For such a scenario, we would need to develop new institutions that could ensure broad-based prosperity even if the labor market's role in distributing income diminishes. Simultaneously, steering technological progress would need to focus more on non-monetary benefits of technology.

Our findings are relevant in multiple domains: for entrepreneurs and innovators seeking to maximize their positive social impact; for unions and work councils representing worker interests; for government funding of research and development; for broader policy frameworks that currently create incentives for labor-saving innovation through higher taxation of labor; and for direct economic incentives that can actively guide innovative efforts to benefit workers.

More generally, technological progress is by definition always a step into the unknown. The more fundamental an innovation, the more unknowns there will be in practice, and the more difficult it will be to apply the proposed policies. Nonetheless, for a great deal of innovative activity, we do have a sense of which factors will benefit and which factors will be hurt. Even if policymakers cannot ascertain this, innovators might be able to. And it may also be possible to guide innovation by committing to implement some of the proposed policies with ex-post measures that are taken once the impact of an innovation is clear.

The future of work will be determined not by blind technological forces but by the choices we make collectively about which technologies to develop and deploy, and how to distribute their benefits. By steering technological progress while simultaneously preparing for potential structural changes in our economy, we can work toward a future where technological advancement benefits all members of society.

References

Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. *Quarterly Journal of Economics*, 113(4):1055–1089.

Acemoglu, D. (2002). Directed technical change. *Review of Economic Studies*, 69(4):781–809.

Acemoglu, D. (2010). When does labor scarcity encourage innovation? *Journal of Political Economy*, 118(6):1037–1078.

Acemoglu, D. (2024). Harms of AI. In Justin Bullock et al., editor, *The Oxford Handbook of AI Governance*, pages 660–706. Oxford University Press.

Acemoglu, D., Manera, A., and Restrepo, P. (2020). Does the US tax code favor automation? *Brookings Papers on Economic Activity*, 2020(1):231–300.

Acemoglu, D. and Restrepo, P. (2018). The race between machine and man: Implications of technology for growth, factor shares and employment. *American Economic Review*, 108(6):1488–1542.

Acemoglu, D. and Restrepo, P. (2019). Artificial intelligence, automation and work. In Agrawal, A. e. a., editor, *The Economics of Artificial Intelligence*. NBER and University of Chicago Press.

Ahmad, S. (1966). On the theory of induced invention. *Economic Journal*, 76(302):344–357.

Atkinson, A. B. and Stiglitz, J. E. (1980). *Lectures on Public Economics*. McGraw-Hill, New York.

Baqaee, D. R. and Farhi, E. (2018). The microeconomic foundations of aggregate production functions. *NBER Working Paper No. 25293*.

Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Oliver, N., Pezoa Rivera, R., Ravindran, B., Riza, H., Rugege, C., Seoighe, C., Sheehan, J., Sheikh, H., Wong, D., and Zeng, Y. (2025). International ai safety report. Technical Report DSIT 2025/001, Department for Science, Innovation and Technology.

Beraja, M. and Zorzi, N. (2025). Inefficient automation. *The Review of Economic Studies*, 92(1):69–96.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.

Costinot, A. and Werning, I. (2023). Robots, trade, and luddism: A sufficient statistic approach to optimal technology regulation. *Review of Economic Studies*, 90(5):2261–2291.

Drandakis, E. M. and Phelps, E. S. (1966). A model of induced invention, growth, and distribution. *Economic Journal*, 76(304):832–840.

Dubber, M. D., Pasquale, F., and Das, S. (2020). *Oxford Handbook of Ethics of AI.* Oxfo.

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

Greenwald, B. C. and Stiglitz, J. E. (2014). *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress.* New York, NY: Columbia University Press.

Guerreiro, J., Rebelo, S., and Teles, P. (2022). Should robots be taxed? *Review of Economic Studies*, 89(1):279–311.

Jones, C. I. (2024). The ai dilemma: Growth versus existential risk. *American Economic Review: Insights*, 6(4):575–590.

Kennedy, C. (1964). Induced bias in innovation and the theory of distribution. *Economic Journal*, 74(295):541–547.

Korinek, A. and Balwit, A. (2024). Aligned with whom? Direct and social goals for AI systems. In Justin Bullock et al., editor, *Oxford Handbook of AI Governance*, pages 65–85. Oxford University Press.

Korinek, A. and Juelfs, M. (2024). Preparing for the (non-existent?) future of work. In Justin Bullock et al., editor, *Oxford Handbook of AI Governance*, pages 746–776. Oxford University Press.

Korinek, A. and Stiglitz, J. E. (2018). Artificial intelligence, worker-replacing technological progress and income distribution. *Working Paper.*

Korinek, A. and Stiglitz, J. E. (2019). Artificial intelligence and its implications for income distribution and unemployment. In Agrawal, A., Gans, J., and Goldfarb, A., editors, *The Economics of Artificial Intelligence: An Agenda.* NBER and University of Chicago Press.

Kurzweil, R. (2005). *The Singularity Is Near.* Viking.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2):132–157.

Samuelson, P. (1965). A theory of induced innovation along kennedy-weisacker lines. *Review of Economics and Statistics*, 47(4):343–356.

Thuemmel, U. (2023). Optimal taxation of robots. *Journal of the European Economic Association*, 21(3):1154–1190.

Zeira, J. (1998). Workers, machines, and economic growth. *Quarterly Journal of Economics*, 113(4):1091–1117.