Redistributing the AI Dividend: Modeling Data as Labor in a Transformative Economy

Sneha Maria Rozario
The Bhawanipur Education Society College

Srishti Dutta
Heritage Institute of Technology

With
BlueDot Impact & Apart Research

Abstract

The economic transformations induced by artificial intelligence (AI) raise pressing distributional concerns. This paper examines the allocation of the AI Dividend—the surplus value generated through AI advancements—and proposes mechanisms to ensure equitable redistribution. Anchoring our analysis in the *Distribution* track, we focus on the Data as Labor (DaL) framework, inspired by Lanier and Weyl, wherein individuals' data contributions are treated as productive labor. We simulate and compare two paradigms: Data as Capital (DaC), in which data is aggregated as corporate capital, and DaL, wherein individuals are compensated for their data. Using comparative economic simulations, we highlight systemic inequalities emerging under DaC and demonstrate the stabilizing potential of DaL structures. We further subdivide the DaL paradigm into two mechanisms: a corporate taxation regime and an individualized data compensation model, proposing a novel formula for micro-level redistribution. The implications of these models for labor markets, inequality, and societal stability are discussed, with a focus on designing incentive-compatible and scalable economic policies. Our findings suggest that recognizing data as labor not only promotes distributive justice but also enhances the long-term sustainability of the AI-driven economy.

1. Introduction

The rapid advancement of artificial intelligence (AI) technologies is generating extraordinary surplus value, termed as the AI Dividend. However, the distribution of this surplus remains highly asymmetric, favoring corporate capital holders over the individuals whose data underpins AI development.

Current economic structures operate predominantly under a Data as Capital (DaC) paradigm, treating data as an owned asset rather than a product of human labor. In contrast, the Data as Labor (DaL) framework, originally proposed by Lanier and Weyl (2018) [1], reimagines data generation as a form of labor deserving recognition and compensation. This paper addresses two key research questions:

1. How does the economic distribution differ between DaC and DaL paradigms?
2. What microeconomic mechanisms can facilitate equitable redistribution of the AI Dividend under a DaL framework?

We focus specifically on the threats posed by rising inequality, wealth concentration, and the invisibilization of digital labor. Our models aim to design scalable redistribution mechanisms that integrate labor market realities and systemic technological change, ensuring that the benefits of AI are broadly shared rather than narrowly captured. By extending Lanier and Weyl's (2018) framework to propose formal redistribution metrics and simulating alternative distribution models, we contribute to building resilient economic systems for an AI-driven future. Our simulations assume a 20-year time horizon (2025–2045), modeling projected Gini coefficients under different economic paradigms. Parameter values were generated synthetically based on theoretical trajectories, and simulations were run in R (RStudio v2024) using the ggplot2 library.

2. Paradigms of Data Ownership: Capital vs. Labor

2.1 The Data as Capital (DaC) Paradigm

The Data as Capital (DaC) model treats data as a proprietary resource, aggregated and monetized by corporations to generate economic returns. In this paradigm, individuals' contributions are unrecognized and uncompensated, reinforcing wealth concentration and network effects. The resulting economic structure favors a small number of dominant firms and exacerbates existing inequalities.

2.2 The Data as Labor (DaL) Paradigm

In contrast, the Data as Labor (DaL) paradigm reconfigures data generation as a form of economic labor. Data contributors are conceptualized as workers whose inputs merit recognition and compensation. This framework seeks to reestablish the link between productivity and wages that DaC structurally severs.

The DaL model fundamentally reimagines data contributors as co-producers of AI value, emphasizing recognition, proportional compensation, and the facilitation of collective bargaining through Mediators of Individual Data (MIDs). Simulation of a DaL economy reveals a redistribution of AI-generated surplus where workers receive compensation streams linked to their data contributions, market entry barriers for AI development are reduced via decentralized data access, and wealth inequality metrics stabilize or decline relative to the DaC baseline.

3. Mechanisms for Operationalizing Data as Labor

3.1 Collective Data Dividend Model

The Collective Data Dividend Model draws inspiration from the California Data Dividend Bill [2] and seeks to establish a broad-based redistribution mechanism anchored in AI and data-related revenue streams. It defines the annual per-person dividend (D) through the formula:

$$D = \frac{R \times \alpha \times \Psi}{P}$$

R represents the total AI and data-related revenue generated within the jurisdiction, α denotes the allocation ratio indicating the percentage of dedicated to the dividend pool, Ψ stands for the Data Labor Multiplier acknowledging the value contribution of human-generated data, and P is the eligible population.

This approach systematizes the recognition of collective data labor without the necessity of tracking individual contributions. Instead, it leverages aggregated economic metrics to ensure that citizens, as implicit data contributors, partake proportionally in AI-driven economic gains. By embedding redistribution into the revenue structure, it aligns technological advancement with societal welfare objectives.

3.2 Individualized Future Economic Value (*FEV)* Model

While collective approaches simplify redistribution, a more granular framework is necessary to recognize differential individual contributions. The Individualized Future Economic Value (*FEV*) model introduces a dynamic, forward-looking formula that accounts for data quality, labor displacement, productivity dynamics, and social adjustment.

The FEV is defined as:

$$FEV = \Psi \times \left\{ \frac{MPC \times \sum(D \times Q)}{N} \times (1+r)^t + \frac{W}{L} \times \lambda \times \eta + \frac{P \times Y}{\pi + \mu} \right\} \times (1-\sigma) \times e^{(\rho - \theta)}$$

Symbol	Represents
Ψ	Data Labor Multiplier – scales all economic outcomes by the recognized importance of human data contributions.
Q	Quality of Data – measures the quality of individual data inputs.
N	Number of Contributors – the total number of contributors providing data.
r	Return Rate – the rate of return on data contribution over time.
t	Time – time factor over which data growth is compounded.
λ	Labor Displacement by AI – the rate at which AI displaces human labor.
W	Wealth – total wealth in the economy.
L	Labor Units – total number of labor units (workers) in the economy.
η	Efficiency – efficiency with which displaced workers are reallocated.
P	Net Productive Output – total productive output in the economy.
Y	Yield – economic yield produced by labor and capital combined.
π	Inflation – the inflation rate affecting the economy.
μ	Model Drift Costs – costs arising from model drift over time, impacting economic predictions.
σ	Social Loss Coefficient – accounts for inefficiencies due to inequality.
θ	Technological Progress – exponential adjustment for technological progress outpacing human adaptability.

Table 1: Symbol Definitions for the FEV Model

This structure ensures that individualized compensation reflects both immediate data contributions and broader systemic economic shifts. It is designed to be adaptable over time as technology, society, and labor markets evolve, ensuring a more resilient and just distribution of the AI Dividend. Upon having calculated the FEV, the individual data dividend can be calculated by : $Di = (\frac{\Psi i}{\Sigma\Psi}) \times FEV$, where, **D_i** is the per capita dividend payment attributable to the individual, representing the monetary compensation derived from their data contributions to the AI economy. **Ψ_i** represents the individual data labor value, quantifying the economic significance of an individual's data input, normalized by the aggregate economic impact of AI-driven processes. **$\Sigma\Psi$** is the aggregate data labor value, representing the total sum of data contributions from the entire population, reflecting the collective input into the AI ecosystem and ***FEV*** symbolises the total Future Economic Value, which represents the projected total economic surplus generated by the AI-driven economy, earmarked for redistribution among the population.

4. Comparative Analysis

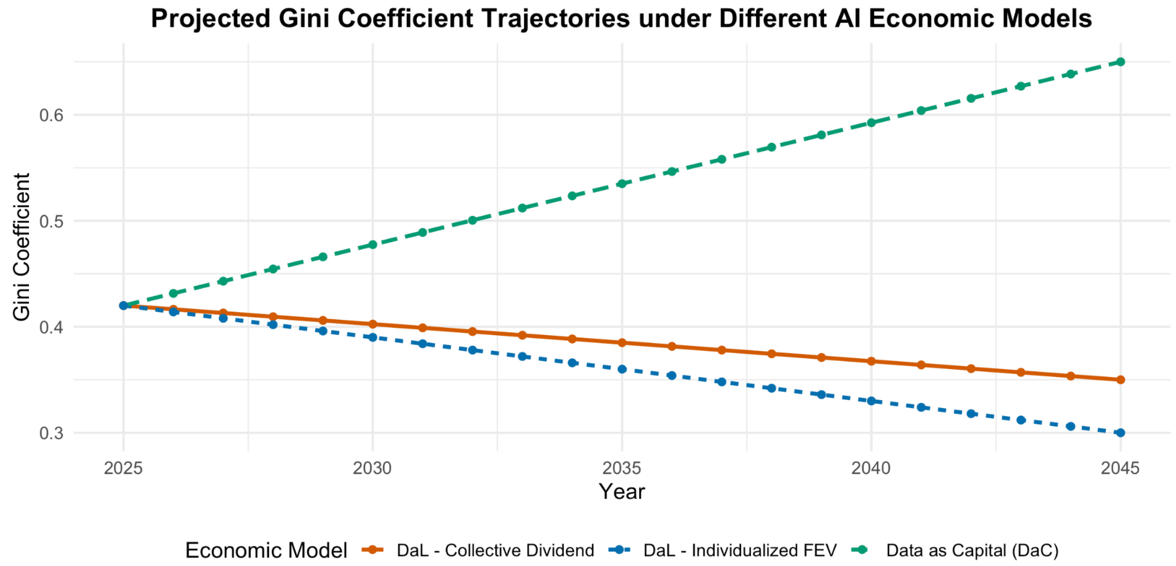

Figure 1: Projected Gini Coefficient Trajectories under Alternative AI Economic Models.

This figure compares projected inequality levels under three paradigms: the Data as Capital (DaC) model, the Data as Labor (DaL) Collective Dividend model, and the DaL Individualized Future Economic Value (FEV) model. Under DaC, inequality rises significantly over time, whereas both DaL models stabilize or reduce inequality, with individualized compensation producing the most equitable outcomes.

Model	**Gini at $t = 0$**	**Gini at $t = 5$**	**Gini at $t = 10$**	**Gini at $t = 15$**	**Gini at $t = 20$**
Data as Capital (DaC)	0.42	0.50	0.57	0.62	0.65
DaL - Collective Dividend	0.42	0.39	0.37	0.36	0.35
DaL - Individualized FEV	0.42	0.36	0.33	0.31	0.30

Table 2: Projected Gini Coefficient Values under Alternative AI Economic Models (2025–2045)

This table presents the simulated Gini coefficients at time points t = 0, 5, 10, 15, 20 for three economic paradigms: Data as Capital (DaC), Data as Labor (DaL) with a Collective Dividend model, and DaL with an Individualized Future Economic Value (*FEV*) model. The results highlight the divergent inequality trajectories between capital-concentrated and labor-inclusive frameworks.

5. Policy Recommendations

To operationalize equitable redistribution of the AI Dividend under the Data as Labor framework, we propose the following policies:

1. **Mandated AI Revenue Contributions:** Require firms generating AI-related revenue above a defined threshold to allocate a fixed percentage (e.g., 10–30%) toward a public dividend fund, scalable by jurisdiction.
2. **Standardized Data Labor Contracts:** Legally recognize data contribution as a form of labor, mandating standardized compensation agreements between individuals and data-using entities.

3. **Dynamic Compensation Metrics:** Implement adaptive formulas, such as the Future Economic Value (FEV) model, to calibrate individual payouts based on data quality, labor market impacts, and productivity shifts.
4. **Social Loss Mitigation Mechanisms:** Establish public investment programs targeting regions and demographics most adversely affected by labor displacement to counteract inequality drag effects.
5. **Technological Adaptability Incentives:** Provide subsidies or tax incentives for firms investing in human reskilling initiatives aligned with AI-driven labor transitions.

Collectively, these policies aim to integrate the economic value of data labor into AI-era wealth distribution, mitigating risks of systemic inequality and social instability.

6. Discussion and Conclusion

This paper explores the distributional challenges of AI-driven economic growth, contrasting the Data as Capital (DaC) and Data as Labor (DaL) paradigms. We find that DaC exacerbates inequality by marginalizing data contributors, while DaL offers a framework for equitable redistribution.

The two models we propose: the Collective Data Dividend, which allocates AI revenues to all contributors, and the Individualized Future Economic Value (FEV) model, which adjusts compensation based on data quality and labor displacement aim to reduce wealth concentration and better integrate labor into the AI economy.

However, the models simplify real-world complexities, particularly around data valuation, labor dynamics, and privacy. Assumptions about uniform data contributions and fixed growth rates may not reflect the true volatility of technological and societal shifts. It does not account for scenarios where AI systems, such as AlphaGo, generate their own training data through self-play. Additionally, privacy and data governance remain key challenges not fully addressed in our framework. Despite these limitations, our work provides a solid foundation for equitable redistribution mechanisms in the AI economy. Further research is necessary to refine these models, particularly with respect to privacy and evolving economic variables.

7. References

[1] Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E. Glen Weyl. 2018. "Should We Treat Data as Labor? Moving beyond 'Free.'" *AEA Papers and Proceedings* 108: 38–42.

[2] Becker, Scott. 2024. *California AI Transparency Act*, SB 942 (2023–2024).

---

8. Appendix

R code used for plotting the graphs given at :
https://gitlab.com/srishti.dutta1111/econoftransformativeai

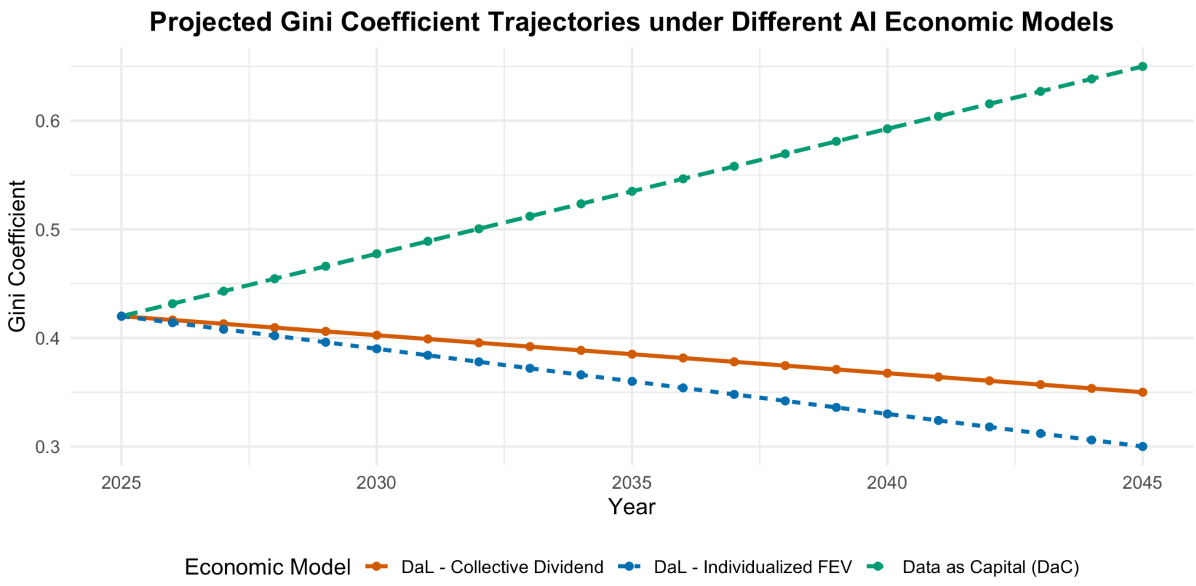

Fig 1: **Projected Gini Coefficient Trajectories under Alternative AI Economic Models.**

This figure compares projected inequality levels under three paradigms: the Data as Capital (DaC) model, the Data as Labor (DaL) Collective Dividend model, and the DaL Individualized Future Economic Value (FEV) model. Under DaC, inequality rises significantly over time, whereas both DaL models stabilize or reduce inequality, with individualized compensation producing the most equitable outcomes.

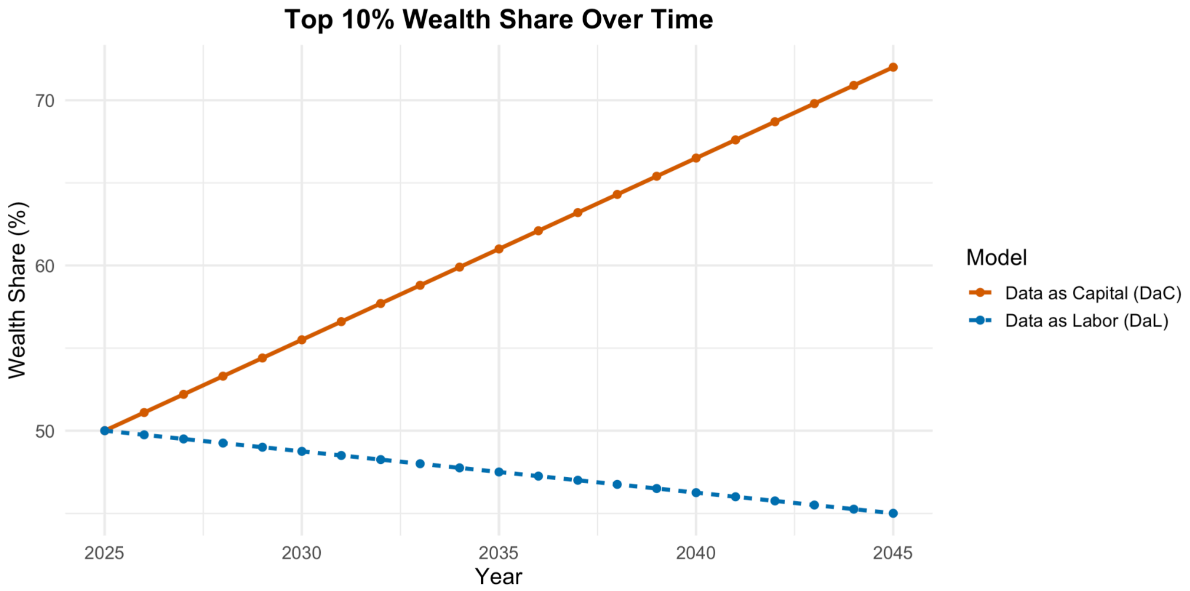

Figure 2: **Projected Wealth Share of the Top 10% under Different AI Economic Models (2025–2045).** *This figure illustrates the divergence in wealth concentration between the Data as Capital (DaC) and Data as Labor (DaL) paradigms. Under DaC, the share of total wealth held by the top 10% rises significantly, whereas under DaL frameworks, redistribution mechanisms stabilize or reduce concentration.*

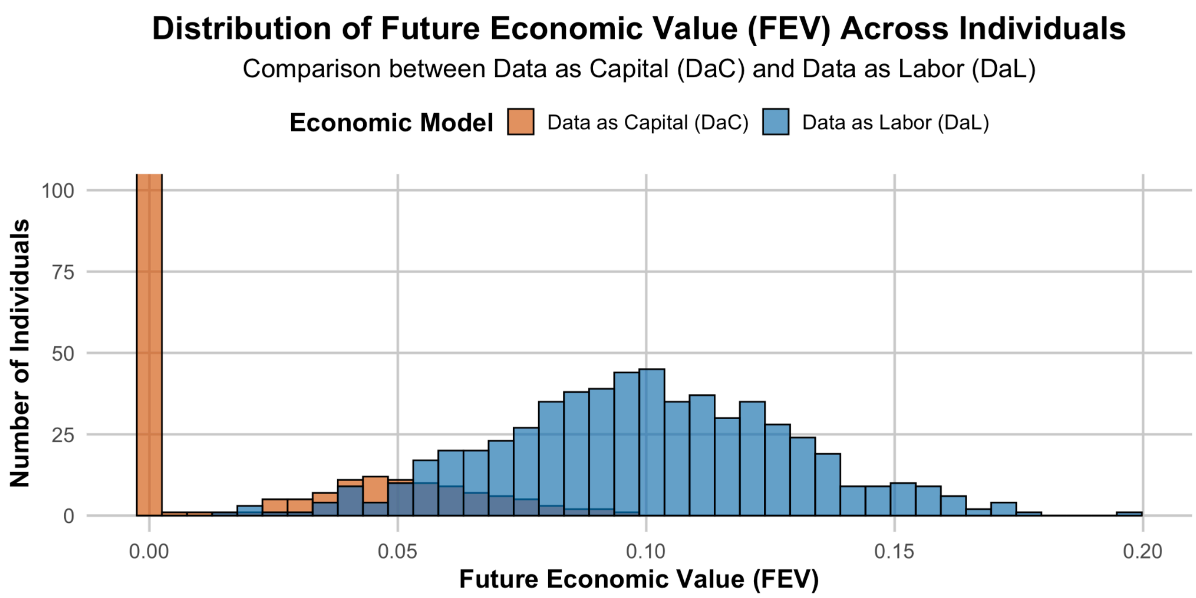

Figure 3: ***Simulated Distribution of Individual Future Economic Value (FEV) under DaC and DaL Models.***
This histogram presents a synthetic simulation of future economic value distribution across individuals. The sample includes 1200 individuals in total (600 modeled under the Data as Capital (DaC) paradigm and 600 under the Data as Labor (DaL) paradigm). Y-axis values represent the absolute number of individuals in each FEV range bin. Under DaC, a majority of individuals derive negligible future economic value, while under DaL models, a broader and more equitable distribution of value is observed.

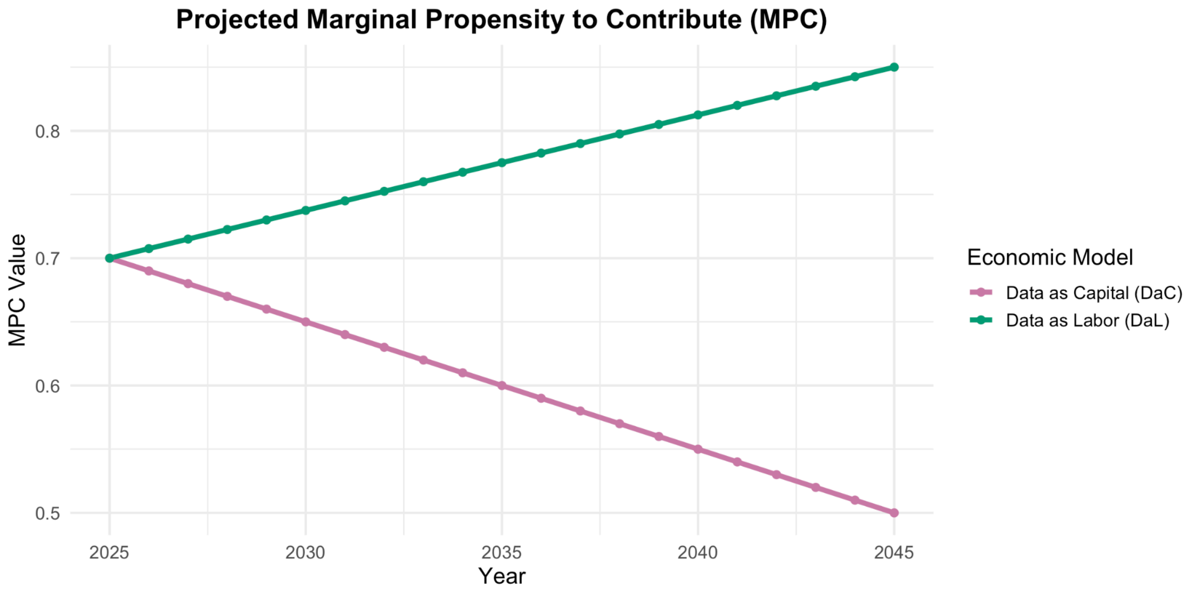

Figure 4: ***Projected Marginal Propensity to Contribute (MPC) Trajectories under DaC and DaL Paradigms (2025–2045).***
The figure models how incentive structures affect individuals' willingness to contribute high-quality data over time. Under DaC, MPC declines due to lack of direct benefits, whereas under DaL, MPC increases due to compensation mechanisms.

<u>Collective Data Dividend Model Calculations.</u>

$\mathfrak{D} = (R \times \alpha \times \Psi) \div P$

Symbol	Meaning
D	Per-person Annual Dividend
R	Total AI/Data-related Revenue generated in California
α	Allocation ratio (percentage of R set aside for the dividend pool)
Ψ	Data Labour Multiplier (value of citizens' data in the AI economy)
P	Eligible population (number of California residents)

<u>Example Simulation for the state of California:</u>

Value	Example
R	\$500 billion AI/Data-related revenue (across tech companies in California)
α	0.15 (15% set aside)
Ψ	0.8 (human data deemed 80% critical)
P	39 million

Now plugging into the formula:

$\mathfrak{D} = (500{,}000{,}000{,}000 \times 0.15 \times 0.8) \div 39{,}000{,}000$

First:

- $500\text{B} \times 0.15 = 75\text{B}$
- $75\text{B} \times 0.8 = 60\text{B}$

Then:

- $60\text{B} \div 39\text{M} \approx \$1{,}538$ per person annually.

So, each Californian would get about $1,538 a year (roughly $128 per month) just for their data contributions, if the AI/data dividend bill passed under these assumptions!

FEV Model Calculations:

$$FEV = \Psi \times \left\{ \frac{MPC \times \sum(D \times Q)}{N} \times (1+r)^t + \frac{W}{L} \times \lambda \times \eta + \frac{P \times Y}{\pi + \mu} \right\} \times (1-\sigma) \times e^{(\rho-\theta)}$$

Plugging in Realistic 2024 US numbers:

Term	Value (2024)	Notes
Ψ (Psi)	0.7	Assume human data accounts for 70% of AI value
MPC (Data)	0.6	60% of people actively contribute usable data
Sum(Data $\times$ Quality)	1,000,000	Arbitrary units; composite measure
N (Contributors)	330,000,000	US population estimate
Rate of Return (r)	0.04 (4%)	Average AI economy growth
Time (t)	5 years	Forecasting 5 years forward
Wealth (W)	$150 trillion	Estimated total US wealth 2024
Labour (L)	160 million	Size of US workforce
Labour Displacement Rate ($\partial L/\partial t$)	0.02	2% of jobs displaced per year
Labour Efficiency (η)	0.5	Only 50% efficiently reskilled
Productivity Growth (P)	0.025 (2.5%)	AI-driven growth
Output (Y)	$26 trillion	US GDP 2024
Inflation (π)	0.03 (3%)	Typical 2024 inflation
Model Decay Cost (μ)	0.01 (1%)	Ongoing AI maintenance cost
Social Loss Coefficient (σ)	0.2	20% loss due to inequality etc.
Tech Progress Rate (ρ)	0.07 (7%)	Rapid AI advancement
Human Adaptability Rate (θ)	0.03 (3%)	Slow reskilling

Step-by-Step Simulation:

First Term: Data Contribution and Compounding

(0.6×(1,000,000÷330,000,000))×(1+0.04)5(0.6 × (1,000,000 ÷ 330,000,000)) × (1 + 0.04)^5

First inside:

- $1{,}000{,}000 \div 330{,}000{,}000 \approx 0.00303$
- $0.6 \times 0.00303 \approx 0.001818$

Compound growth for 5 years at 4%:

- $(1.04)^5 \approx 1.2167$

Thus:

- $0.001818 \times 1.2167 \approx 0.002213$

Second Term: Labour Displacement and Wealth Efficiency

$$(150{,}000{,}000{,}000{,}000 \div 160{,}000{,}000) \times 0.02 \times 0.5$$

First inside:

- $150 \text{ trillion} \div 160 \text{ million} \approx 937{,}500$

Then:

- $937{,}500 \times 0.02 \times 0.5 \approx 9{,}375$

Third Term: Productivity vs Inflation and Decay

$$(0.025 \times 26{,}000{,}000{,}000{,}000) \div (0.03 + 0.01)$$

First inside:

- $0.025 \times 26 \text{ trillion} = 650 \text{ billion}$

Then denominator:

- $0.03 + 0.01 = 0.04$

Thus:

- $650 \text{ billion} \div 0.04 \approx 16.25 \text{ trillion}$

Sum of three major terms:

$$0.002213 + 9{,}375 + 16{,}250{,}000{,}000{,}000$$

Given how tiny the first two are compared to trillions, for practical purposes:

- Major driver is 16.25 trillion.

Social Adjustment: $(1 - \sigma) = 1 - 0.2 = \mathbf{0.8}$

Exponential Growth Correction:
$e^{(\rho - \theta)} = e^{(0.07 - 0.03)} = e^{(0.04)} \approx 1.0408$

Now final FEV calculation:

$\text{Future Economic Value} = 0.7 \times 16.25 \text{ trillion} \times 0.8 \times 1.0408$

First inside:

- $16.25 \text{ trillion} \times 0.8 \approx 13 \text{ trillion}$
- $13 \text{ trillion} \times 1.0408 \approx 13.5304 \text{ trillion}$

Then:

- $0.7 \times 13.5304 \text{ trillion} \approx 9.47128 \text{ trillion}$

Final Simulation Result:

> Future Economic Value (FEV) ≈ $9.47 trillion increase over 5 years, driven by human data contributions, AI productivity, labour adaptation, and social management!

Interpretation:

- Despite inflation, labour displacement, and model decay,

- If we properly value human data (Psi), retrain workers efficiently, and manage inequality, The United States could add nearly $9.5 trillion to its economy by 2029, just from the AI-data ecosystem.

Individual Dividend System Using Above Simulation

From earlier:

- Total Future Economic Value ($\mathfrak{F}$EV) $\approx$ $9.47 trillion over 5 years
- Population (P) $\approx$ 39 million Californians
- Ψ (Psi for human data value) $= 0.7$

Now:

Total Dividend Pool:

Assuming $\alpha = 15\%$ of R is set aside (as above), and $\Psi = 0.7$,
then Dividend Pool =

Dividend Pool=9.47 trillion$\times$0.15$\times$0.7

First inside:

- 9.47 trillion $\times$ 0.15 = 1.4205 trillion
- 1.4205 trillion $\times$ 0.7 $\approx$ 994.35 billion dollars

Thus:

Dividend Pool $\approx$ $994.35 billion over 5 years

Base $\mathfrak{D}$: Per-person Dividend Over 5 Years

$D=\frac{994{,}350{,}000}{39{,}000{,}000}\approx 25{,}497$

Thus:

$\mathfrak{D}$ $\approx$ $25,497 per person over 5 years,
or about $5,099 per year per person, *before individual adjustments*.

Now Adding Individual Adjustment by Data Labour (Ψ_i)

The formula again:

$D_i = \frac{\Psi_i}{\Sigma\Psi} \times D$

Assume:

- $\sum\Psi$ (total data labor across population) = 39,000,000 units (1 unit per person on average)
- Individual Ψ_i examples:

Type of Contributor	Ψ_i Value	Calculation	Annual Dividend (D_i)
Minimal data sharer	0.5	$(0.5 \div 39,000,000) \times 5,099$	$\approx$ $0.065
Average citizen	1	$(1 \div 39,000,000) \times 5,099$	$\approx$ $0.13
Heavy contributor	5	$(5 \div 39,000,000) \times 5,099$	$\approx$ $0.65
Super contributor (e.g., influencer)	50	$(50 \div 39,000,000) \times 5,099$	$\approx$ $6.53