web3 foundation

2026

TECHNICAL METHODOLOGY

The Hidden Price of Free: What Your Data Is Really Worth

Personal Data Annual Value (PDAV)

Calculation architecture, notation, data pipeline, uncertainty treatment and audit model

A Web3 Foundation White Paper

Publication Date
25 May 2026

Scope

This paper specifies the technical methodology used to calculate Personal Data Annual Value (PDAV) and the associated strict-attribution floor metric, PDSEV-AI. It covers input schema, symbol definitions, firm qualification, archetype parameterisation, channel attribution, population allocation, regional scaling, scenario assembly, output persistence and known methodological limitations.

The scope is limited to the calculation engine and its audit trail. It specifies how inputs are transformed into outputs, how assumptions are parameterised and how uncertainty is presented.

1. Metric stack and calculation purpose

The metric family evolved through six stages, each answering a narrower or more economically useful question than the previous version. The sequence is retained because intermediate metrics provide useful audit checks even when the published headline is PDAV.

Metric	Technical role	Principal calculation boundary
DEAI	Structural index of a firm's data-extraction architecture across aggregation, reusability, feedback loops and opacity.	Index only; no revenue denominator.
HDAF	Human-Data Attributable Flow. Firm-level annual value obtained by reducing channel value to the human-data-attributable component and allocating it across populations.	Firm-level dollar flow, not per person.
PUEV	Per-User Extracted Value. The active-user portion of HDAF divided by active users.	Active-user denominator only.
PDSEV	Personal Data Subject Extracted Value. HDAF divided by the deduplicated set of data subjects touched by the firm.	All-subject denominator, without the strict AI-attribution split.
PDSEV-AI	Strict-attribution floor. The human-data-attributable flow is further restricted by the AI attribution share ψ^attr.	Causal within-year booked-revenue attribution.
PDAV	Headline annual value. Regional ecosystem calculation using booked channel value multiplied by the AI-era leverage multiplier λ^lev.	Ecosystem-level annual value, divided by a capped regional active-user population.
PDLV	Personal Data Lifetime Value. Inflation-indexed lifetime extension of PDAV over the model horizon. (Lifetime restatement of PDAV)	Lifetime aggregation over 60 years, reported in two anchors side-by-side: nominal forward-projected at the regional CPI (AF^{nom}), and real / today's-$ (AF^{real} = H). Neither is risk-discounted.

PDLV, Personal Data Lifetime Value, is the lifetime extension of PDAV across the model horizon, reported in two anchors side-by-side: nominal forward-projected at the regional CPI assumption, and real / today's purchasing power.

2. Notation and stored inputs

All firm-level values are stored by scenario band. The low, central and high bands are used by the Conservative, Central and Expansive scenarios unless outcome sorting is applied at display time. Monetary fields are stored in billions of dollars; population fields are stored in millions of persons. The final per-person outputs multiply by 1,000 because billions divided by millions equals thousands of dollars.

Symbol	Meaning
f	Firm in the qualifying cohort.
c	Monetisation channel: advertising, subscription, API, enterprise, marketplace, licensing, cost savings or other. The API, enterprise and cost-savings channels are AI-flagged and so consume γ_model · η_human in place of δ · η; the cost-savings channel additionally stores avoided cost rather than booked revenue.
p	Tracked population: active users, non-user data subjects, contributors, public-corpus subjects, inferred persons or paying customers. Only the first five enter the data-subject denominator S; paying customers is a subset of active users used solely as a denominator for revenue-per-paying-customer ratios on the subscription, API and enterprise channels.
r	Region used for reporting and denominator capping: Global, USA, North America, Europe, Rest of World.
s	Scenario band: Conservative, Central or Expansive.
R_{f,c}^s	Annual channel revenue — or avoided cost for the cost-saving channel — in billions of dollars.
δ_{f,c}^s	Channel-level data-dependence share. Applied as R · δ · η for non-AI channels to derive the data-attributable channel value used by HDAF, PDSEV and PDSEV-AI.
η_{f,c}^s	Person-level human-data attribution share, applied alongside δ for non-AI channels.
γ_{model, f,c}^s	Model-attributable share. Used in place of δ on AI-flagged channels where contract revenue includes hosting, integration or consulting bundles that are not data-driven.
η_{human, f,c}^s	Human-data attribution share used in place of η on AI-flagged channels, so the data-attributable channel value becomes R · γ_model · η_human.
ψ^attr_{f,c}^s	AI attribution share — within-firm fraction of channel-c revenue causally attributable to ML use of human data. Range [0, 1]. Consumed by PDSEV-AI as a split operator: R · δ · η · ψ^attr. Stored as a single per-channel per-firm value; λ^lev below reads the same stored value under the calibration identity λ^lev := 1 + ψ^attr.

Symbol	Meaning
λ^lev_{f,c}^s	AI-era leverage multiplier. Range [1, 2]. Consumed by PDAV as an amplify operator: R · λ^lev. Not independently stored — derived as 1 + ψ^attr at read time.
revShare_{f,r}^s	Fraction of firm f's revenue attributable to region r, scenario s. Sourced from 10-K disclosures for the major US-listed firms; an internet-population-weighted default applies elsewhere.
subjShare_{f,r}^s	Fraction of firm f's active users or subjects attributable to region r, scenario s.
τ_f	Territorial allocation weight; default 1.0. Applied as a per-firm multiplier inside the cohort sum. Scenario-invariant in the current model.
U_f^s	Firm f's active-user count in scenario s, in millions.
cap_r	Regional internet-using-population cap in millions: Global 6,000; USA 322.064; North America 470.796; Europe 484.7; Rest of World 4,565. Sourced primarily from World Bank 2024 (23/25 also)).
N_{f,p}^s	Population count for firm f, population p, scenario s, in millions.
κ_{f,c,p}^s	Intensity weight connecting channel c to population p. Captures how much each population feeds each monetisation channel.
a_{f,c,p}^s	Normalised channel-population allocation share, a_{f,c,p} = (κ_{f,c,p} · N_p) / Σ_q (κ_{f,c,q} · N_q). Row sums to 1.0 by construction so allocation does not double-count across populations.
S_f^s	Deduplicated count of data subjects for firm f, scenario s: S = U + non-user data subjects + contributors + public-corpus subjects + inferred persons − overlap. Paying customers are not included. Values below 1 million are treated as not computable to avoid divide-by-tiny artefacts.
r_{infl, r}	Central annual CPI assumption for region r. North America 2.30%, Europe 2.10%, Global 3.25%. Resolved per (cohort, region) from the per-region inflation table, with fallback to a singleton default, finally 0. No Greek symbol is assigned because π is already in use for the paying-user share.
H	Lifetime projection horizon. Configured as 60.3 (UN DESA 2024 global life expectancy at birth, 73.3, minus the COPPA child-consent threshold, 13). Rounded to the nearest integer at the resolver boundary before the engine applies it. Fractional final-year prorating is not implemented; H is treated as a pure integer year count. Operative value: 60.
AF_r^{nom}	Nominal lifetime aggregation multiplier (forward-projected, undeflated, future dollars). AF_r^{nom} = Σ_{t=0..H−1} (1 + r_{infl,r})^t = ((1 + r_{infl,r})^H − 1) / r_{infl,r}. Live values: USA / NA 126.66×, Europe 118.08×, Global / RoW 178.89×.
AF_r^{real}	Real (today's purchasing power) lifetime aggregation multiplier. AF_r^{real} = H under the assumption that revenue grows at exactly r_{infl,r}, so deflating each future year by r_{infl,r} cancels the growth and leaves a flat repetition of today's annual flow. Live value: 60× in every region.

4. Channel and population schema

For every firm, the model stores low, central and high estimates for channel revenues, population counts, overlap fields, regional shares and parameter overrides. Each value carries provenance metadata, citation fields, evidence summaries and confidence information.

Monetisation channel	Technical treatment
Advertising	Booked advertising revenue attributable to targeting, ranking, auction optimisation or similar commercial mechanisms.
Subscription	User or customer subscription revenue where the product is materially improved by personal or person-linked data.
API	Developer or model-access revenue, adjusted for model/data dependence where applicable.
Enterprise	Contract and enterprise revenue, adjusted for bundled implementation, hosting or consulting components where applicable.
Marketplace	Transaction, commission or marketplace revenue dependent on matching, ranking, trust or behavioural data.
Licensing	Data, model, rights or analytic licensing revenue.
Cost savings	Avoided cost stored directly as value rather than revenue.
Other	Residual monetisation channel with documented evidence and parameters.
Tracked population	Definition used in the model
Active users	Persons who directly use the firm's product or service.
Non-user data subjects	Persons profiled, inferred about or captured in datasets without being active users.
Contributors	Persons who contribute content, ratings, labels, posts, prompts, code, media or other data inputs.
Public-corpus subjects	Persons represented in public or web-scraped corpora used by the firm.
Inferred persons	Persons represented indirectly through household, graph, lookalike or probabilistic inference.
Paying customers	Persons or accounts paying directly; deduplicated against active users where applicable.

5. PDAV headline calculation

PDAV is the regional headline calculation. It takes each firm-channel value, applies the AI-era leverage multiplier λ^lev, scales the result to the reporting region, sums across the qualifying cohort, and divides by a capped regional active-user population.

$$PDAV_r^s = 1000 \times [\ \Sigma_f \Sigma_c R_{f,c}^s \times \lambda^{lev}_{f,c}^s \times revShare_{f,r}^s\ ] / min[\ \Sigma_f \tau_f \times U_f^s \times subjShare_{f,r}^s, cap_r\ ]$$

The numerator is an ecosystem-level annual value in billions of dollars: each firm contributes its booked channel revenue (or avoided cost on the cost-savings channel) amplified by λ^lev and scaled by its share of the reporting region's revenue. The denominator is a regional active-user population in millions, capped at the regional internet-using population cap_r so the same person is not counted repeatedly across firms. The cap binds almost immediately for any cohort that includes two or more large platforms; once it binds, adding further firms strictly raises the headline because the numerator grows while the denominator is held at population reality. The multiplication by 1,000 converts $bn / M-persons into dollars per person per year.

Operator and the two-symbol calibration

The PDAV operator is R × λ^lev. The leverage multiplier λ^lev lies in [1, 2] and is read at the operator as λ^lev := 1 + ψ^attr, where ψ^attr is the within-firm AI attribution share consumed by PDSEV-AI. The two symbols share one stored value per channel per firm; the operator pivot — split for PDSEV-AI, amplify for PDAV — is the only thing that differs.

Because λ^lev already contains the booked-revenue base inside its +1 term, PDAV is not calculated as R × (1 + λ^lev). That form would add the booked base a second time and produce the double-count that the operator is specifically designed to avoid. The indirect leverage layer can be read as R × (λ^lev − 1) = R × ψ^attr, and the total channel contribution is R × λ^lev.

PDAV stream	Expression	Interpretation
Booked value	R	Observed current-year channel revenue, or avoided cost on the cost-savings channel.
Data leverage layer	R × (λ^lev − 1)	First-order proxy for the AI-era data leverage that sits outside current-year booked revenue — option / training-stock value, data network effects, scaling-law productivity and surplus-capture potential. Calibrated using the same empirical anchor as ψ^attr because that is the best evidence the literature currently supports.
Total PDAV channel value	R × λ^lev	Booked value plus the calibrated leverage layer. The sum of two distinct economic flows.

6. PDSEV-AI strict-attribution calculation

PDSEV-AI is the rigorous floor metric. It uses a revenue-splitting operator rather than a leverage multiplier. The calculation restricts channel value to the portion attributable to data-derived outputs, human or person-level data and AI or machine-learning systems.

$$\mathbf{HDAF\_AI(f,p,s) = \Sigma_c\ R_{f,c}^s \times \delta_{f,c}^s \times \eta_{f,c}^s \times \psi^{attr}_{f,c}^s \times a_{f,c,p}^s}$$
$$\mathbf{PDSEV\_AI(f,s) = 1000 \times [\Sigma_p\ HDAF\_AI(f,p,s)] / S_f^s}$$

For ordinary non-AI channels, the strict human-data-attributable flow is built from $R \times \delta \times \eta$. For AI-flagged channels (API and enterprise), the implementation uses model-specific shares γ_{model} and η_{human} in place of δ and η, isolating the model-attributable and human-data-attributable portions of bundles that also include hosting, integration and consulting. For the cost-savings channel, value is stored as avoided cost and is already attributable by construction. In every case the same AI attribution share ψ^{attr} applies on top of the channel core to isolate the ML-derived portion.

The data-subject denominator S is a deduplicated count of persons whose data the firm uses:

$$\mathbf{S_f = active\_users + non\_user\_data\_subjects + contributors + public\_corpus\_subjects + inferred\_persons - overlap}$$

Overlap fields prevent the four non-user populations, non-user data subjects, contributors, public-corpus subjects and inferred persons, from being counted a second time when those persons are also active users of the firm.

7. AI attribution share and leverage multiplier

The methodology separates the strict attribution share from the headline leverage multiplier. The same empirical evidence can calibrate both, but they are different symbols because they are used in different operators and answer different questions.

Symbol	Where used	Operator	Role
$\psi^{attr}_{f,c}^s$	PDSEV-AI only	$R \times \delta \times \eta \times \psi^{attr}$	Within-firm causal attribution share. Range: [0, 1].
$\lambda^{lev}_{f,c}^s$	PDAV only	$R \times \lambda^{lev}$	AI-era leverage multiplier. Range: [1, 2].

The symbols are tied together at exactly one calibration point:

$$\mathbf{\lambda^{lev}_{f,c}^s := 1 + \psi^{attr}_{f,c}^s}$$

The implementation stores a single AI-attribution parameter per channel per firm. In the strict-attribution pathway that stored value is read as ψ^attr. In the PDAV pathway it is converted once into λ^lev (= 1 + ψ^attr) before the PDAV formula is evaluated. After that conversion, PDAV uses λ^lev only.

The advertising defaults are anchored to empirical evidence from Apple's App Tracking Transparency natural experiment. The 2024 Aridor and Che working paper reports material degradation in Facebook advertiser targeting performance after ATT. A follow-on paper by Aridor, Che, Hollenbeck, Kaiser and McCarthy reports firm-side advertising-performance and revenue effects. These sources support the use of a material AI or machine-learning attribution share for ad-funded platforms, while firm-specific overrides are used where better evidence is available [1,2].

The broader leverage interpretation is supported by literature treating data as an asset, examining data-related network effects, documenting scaling relationships between model performance and data, studying human-feedback-based model improvement, analysing web-scraped personal data in training corpora and measuring consumer surplus from free digital goods [3-9].

Central ψ^attr defaults: commercial channels

Archetype	Advertising	Subscription	API	Enterprise
Ad-funded platform	0.55	0.35	1	0.95
AI-first subscription/API	0.45	0.98	1	0.95
Data broker	0.45	0.3	1	0.95
Hardware-bundled	0.45	0.35	1	0.95

For PDAV, the corresponding λ^lev value equals 1 plus the table value. For example, a ψ^attr value of 0.55 maps to λ^lev = 1.55.

Central ψ^attr defaults: ancillary channels

Archetype	Advertising	Subscription	API	Enterprise
Ad-funded platform	0.45	0.25	1	0.35
AI-first subscription/API	0.4	0.25	1	0.55
Data broker	0.4	0.55	1	0.35
Hardware-bundled	0.4	0.25	1	0.3

Worked channel example

Assume $100 of advertising revenue, δ = 1, η = 1 and ψ^attr = 0.55. The calibration identity gives λ^lev = 1.55.

Metric	Calculation	Interpretation
PDSEV-AI	$100 × 1 × 1 × 0.55 = $55	Approximately $55 of booked annual advertising revenue is treated as AI-attributable under the strict within-year attribution interpretation.
PDAV	$100 × 1.55 = $155	The channel contribution is interpreted as $100 of booked value plus a $55 data-leverage layer outside the current-year booked revenue stream.
Metric	Calculation	Interpretation
PDSEV-AI	$100 × 1 × 1 × 0.55 = $55	Approximately $55 of booked annual advertising revenue is treated as AI-attributable under the strict within-year attribution interpretation.

ψ^attr is structurally the single most consequential parameter family in the model.

It appears directly in PDSEV-AI through the split operator R × δ × η × ψ^attr, and in PDAV through the leverage identity λ^lev := 1 + ψ^attr.

The AI gather attempts a firm-specific value of ψ^attr for every cohort firm. Where public disclosure is sufficient, that firm-specific value takes precedence over the archetype default.

Where disclosure is thin, the parameter falls back to the archetype default. Whether a firm gets its own value or the default depends on disclosure, not on firm size.

Across the 129-firm cohort the AI gather lands a firm-specific ψ^attr value on ~96% of firms depending on channel, summarised in the table below.

The largest concentration of archetype-default fall-back is on api (12 of 129 firms) and lic (11 of 129), both small contributors to the cohort headline. On the four largest channels by cohort HDAF (sub, ad, ent, mkt) the firm-specific coverage is between 96% and 99%.

ψ key	Firm-specific override	Archetype default	Total
ai_uplift_ad	124	5	129
ai_uplift_sub	125	4	129
ai_uplift_api	117	12	129
ai_uplift_ent	127	2	129
ai_uplift_mkt	128	1	129

ψ key	Firm-specific override	Archetype default	Total
ai_uplift_lic	119	10	129
ai_uplift_cs	128	1	129
ai_uplift_other	128	1	129

The following sensitivity table quantifies how the USA Central headline responds to a ±0.10 shift in ψ^attr per channel, applied uniformly across all firms in the cohort (regardless of whether their current value is a firm-specific override or an archetype default).

Each row tests one parameter at a time: the named ψ^attr channel is shifted; every other input, revenue figures, population counts, all remaining parameters, stays at its Central-scenario value.

This isolates the headline impact of that single ψ^attr channel. Results are shown per channel and as a joint shift across all eight channels (the cohort-wide swing). The shifts are linear in ψ^attr and therefore symmetric.

ψ^attr shift	Cohort HDAF share	PDSEV-AI floor ($/yr)	Δ vs baseline	PDAV headline ($/yr)	Δ vs baseline
Baseline (no shift)	100%	236.52	-	6,563.08	-
ψ^attr_sub −0.10	43%	159.14	−77.38	6,210.32	−352.76
ψ^attr_sub +0.10	43%	313.9	77.38	6,915.84	352.76
ψ^attr_ad −0.10	24%	193.67	−42.85	6,484.08	−79.00
ψ^attr_ad +0.10	24%	279.37	42.85	6,642.08	79
ψ^attr_mkt −0.10	13%	212.35	−24.17	6,506.36	−56.72
ψ^attr_mkt +0.10	13%	260.69	24.17	6,619.80	56.72
ψ^attr_ent −0.10	11%	216.06	−20.46	6,501.51	−61.57
ψ^attr_ent +0.10	11%	256.98	20.46	6,624.65	61.57
ψ^attr_api −0.10	5%	228.2	−8.32	6,538.60	−24.48
ψ^attr_api +0.10	5%	244.84	8.32	6,587.56	24.48
ψ^attr_cs −0.10	2%	232.32	−4.20	6,558.88	−4.20
ψ^attr_cs +0.10	2%	240.72	4.2	6,567.28	4.2

ψ^attr shift	Cohort HDAF share	PDSEV-AI floor ($/yr)	Δ vs baseline	PDAV headline ($/yr)	Δ vs baseline
ψ^attr_other −0.10	1%	234.85	−1.67	6,545.81	−17.27
ψ^attr_other +0.10	1%	238.19	1.67	6,580.35	17.27
ψ^attr_lic −0.10	<1%	236.44	−0.08	6,562.92	−0.16
ψ^attr_lic +0.10	<1%	236.6	0.08	6,563.24	0.16
Joint −0.10 across all 8 channels	100%	57.37	−179.15	5,966.93	−596.15
Joint +0.10 across all 8 channels	100%	415.67	179.15	7,159.23	596.15

The PDAV headline is robust to a ±0.10 shift on any single ψ^attr channel parameter.

The largest single-channel sensitivity is on ψ^attr_sub (±5.4%, reflecting subscription's 43% share of cohort HDAF). The joint ±0.10 shift across all eight channels - the cohort-wide swing - moves the headline by ±9.1%, well inside the published Conservative-to-Expansive scenario band ($4,815.71 – $8,526.93).

PDSEV-AI is materially more sensitive in proportional terms: ±32.7% on ψ^attr_sub and ±75.7% on the cohort-wide ±0.10 swing. The greater proportional sensitivity reflects the multiplicative operator R × δ × η × ψ^attr in PDSEV-AI versus the additive R × (1 + ψ^attr) in PDAV and reinforces PDSEV-AI's role as a strict-attribution floor rather than a precise point estimate.

8. Regional scaling and cohort population caps

Regional scaling ensures that a regional headline reports regional monetisation against a relevant regional population. Each firm-channel contribution is multiplied by revShare for the selected region. The denominator uses active-user counts adjusted by subjShare and τ, then applies a population cap.

$$\textbf{Denominator}(r,s) = \min[\Sigma_f \tau_f \times U_f^s \times \text{subjShare}_{f,r}^s, \text{cap}_r]$$

The cap is the structural mechanism that prevents cross-firm denominator inflation. Without the cap, a cohort containing several large platforms would count the same person repeatedly. With the cap, denominator growth stops at the regional internet-using population, while additional qualifying firms continue to add their regional value to the numerator.

Region	Model cap parameter
Global	6,000 million
USA	322.064 million
North America	470.796 million
Europe	484.7 million
Rest of World	4,565 million

Cap values are editable model parameters. The source series used for cap governance is the World Bank indicator for individuals using the Internet, which is sourced from the International Telecommunication Union [10].

Population allocation

Channel value is allocated across the tracked populations using a count-weighted intensity matrix. The allocation share is:

$$a_{f,c,p}^s = [\kappa_{f,c,p}^s \times N_{f,p}^s] / [\Sigma_q \kappa_{f,c,q}^s \times N_{f,q}^s]$$

This construction guarantees that each channel-allocation row sums to 1.0. It prevents channel value from being counted more than once across populations while allowing the analyst to express that some populations contribute more intensively to a given channel than others.

9. Data-gathering and deterministic calculation pipeline

The operational pipeline has four stages: qualification, archetype classification, per-firm research and deterministic calculation.

1. Qualification applies the seven-criterion rubric and records the pass/fail decision.
2. Archetype classification scores each firm against the four archetype rubrics and persists the full scoreboard.
3. Per-firm research collects channel revenues, population counts, overlap fields, parameter overrides, regional revenue shares and regional subject shares.
4. The calculation engine resolves parameters, builds denominators, iterates through channels, computes strict attribution and headline value, applies population allocation, aggregates outputs and persists results with an input hash.

Research is performed by an AI gather endpoint with web search enabled. It returns proposed low, central and high values, evidence summaries, key findings, confidence ratings and citations. Values are written to the model store with provenance fields.

Coverage item	Status

Table 1: Cohort composition by AI-gather source category

Source category	Firms	Share
AI-gathered from primary regulatory disclosure	86	66.70%
AI-gathered from secondary sources only	43	33.30%
Total	129	100%

produce the same outputs. A calculation run records the triggering user, timestamp, scenario configuration, warnings, status and inputs hash. Universe snapshots are persisted one row per model, cohort, region, scenario and period year.

All inputs in this model are produced by the same AI gather pipeline. What varies across the cohort is the source the gathering draws from. Firms covered by primary regulatory disclosure - the firm's own 10-K, 20-F or equivalent audited annual report - have inputs drawn from audited filings. The remaining firms have no such regulatory filing covering them, and the AI gather must draw on secondary sources such as:

- investor-day briefings
- fundraising disclosures
- journalist reporting
- industry research
- similar public material

Both pathways carry the same provenance, citation and confidence metadata, but the upstream source quality differs.

Table 1: Cohort composition by AI-gather source category

Source category	Firms	Share
AI-gathered from primary regulatory disclosure	86	66.70%
AI-gathered from secondary sources only	43	33.30%
Total	129	100%

Table 2: Composition by archetype

Archetype	Total	Primary regulatory disclosure	Secondary sources only
Ad-funded platform	31	30	1
AI-first subscription / API	65	25	40
Data broker	19	17	2
Hardware-bundled	14	14	0
Total	129	86	43

10. Reported outputs

The methodology reports PDAV and PDSEV-AI side by side because they answer different technical questions. PDSEV-AI is the strict within-year, within-firm, AI-attributable floor. PDAV is the broader regional ecosystem metric using the leverage multiplier and the capped active-user denominator.

Metric	Central USA output	Interpretation
PDSEV-AI floor	$236.52/year	Strict AI-attributable, human-data-attributable value divided by the broad data-subject denominator.
PDAV headline	$6,563.08/year	Regional ecosystem value calculated with λ^lev and a capped regional active-user denominator.
PDSEV-AI lifetime floor	$29,958.54 over 60 years	Inflation-indexed, undiscounted lifetime extension of the strict annual floor using the USA CPI assumption.
PDSEV-AI lifetime floor (real, today's $)	$14,191.27 over 60 years	Same horizon at the real anchor (= annual × H).
PDLV headline (nominal)	$831,300.64 over 60 years	Inflation-indexed, undiscounted lifetime extension of the PDAV headline at the USA CPI assumption.
PDLV headline (real, today's $)	$393,784.73 over 60 years	Same horizon at the real anchor (= annual × H).

The difference between the two outputs follows from two design choices: PDSEV-AI uses the splitting operator δ × η × ψ^attr and the broad data-subject denominator; PDAV uses the multiplier λ^lev and the capped regional active-user denominator.

11. Lifetime-value conversion and inflation assumptions

PDLV converts an annual PDAV value into a lifetime value over the model horizon, reported in two anchors side-by-side: nominal forward-projected at the regional CPI assumption, and real / today's purchasing power. It is a time-aggregation layer applied after the annual per-person value has been calculated.

The primary paid benchmark for the inflation assumptions is Consensus Economics' long-term forecast suite. Public cross-checks are taken from the IMF World Economic Outlook, OECD inflation forecasts, central-bank inflation targets and long-dated inflation-market pricing where available [11-20].

The central CPI assumptions used in the current lifetime run are:

Region	Central annual CPI assumption	Lifetime horizon	Application
USA	2.30%	60 years	USA PDLV and USA lifetime strict-floor outputs.
Europe	2.10%	60 years	Europe PDLV outputs.
Global	3.25%	60 years	Global PDLV outputs.

North America inherits the USA rate (2.30%); Rest of World inherits the Global rate (3.25%). The full per-region multiplier table is shown below.

Let PDAV_0(r,s) denote the current-year PDAV value for region r and scenario s. Let r_{infl,r} denote the regional annual CPI assumption and H the integer lifetime horizon (H = 60, rounded from a configured 60.3 at the resolver boundary; fractional-year prorating is not applied).

$$AF\_r^{nom} = ((1 + r_{infl,r})^H - 1) / r_{infl,r}, \quad \text{for } r_{infl,r} \neq 0$$
$$AF\_r^{nom} = H, \quad \text{for } r_{infl,r} = 0$$
$$AF\_r^{real} = H$$

$$PDLV\_nominal(r,s) = PDAV\_0(r,s) \times AF\_r^{nom}$$
$$PDLV\_real(r,s) = PDAV\_0(r,s) \times AF\_r^{real}$$
$$PDSEV\_AI\_LTV\_nominal(r,s) = PDSEV\_AI\_0(r,s) \times AF\_r^{nom}$$
$$PDSEV\_AI\_LTV\_real(r,s) = PDSEV\_AI\_0(r,s) \times AF\_r^{real}$$

The horizon is treated as integer 60 years in line with the engine's resolver-boundary rounding. The nominal calculation forward-projects the annual flow at the regional CPI assumption and sums

in future dollars without deflation. The real calculation collapses to annual × H under the engine's assumption that revenue grows at exactly the CPI rate. Neither is discounted to present value unless a separate discount-rate layer is explicitly applied.

Region	r_{infl,r}	AF_r^{nom} (60y)	AF_r^{real} (60y)	Terminal CPI multiplier (60y)
USA	2.30%	126.66×	60×	3.913×
North America	2.30%	126.66×	60×	3.913×
Europe	2.10%	118.08×	60×	3.479×
Global	3.25%	178.89×	60×	6.815×
Rest of World	3.25%	178.89×	60×	6.815×

Using the central USA annual outputs, the inflation-indexed 60-year conversion gives $29,958.54 for the PDSEV-AI lifetime floor (nominal) and $831,300.64 for the PDLV headline (nominal). The corresponding real / today's-$ values are $14,191.27 and $393,784.73 respectively.

12. Scenario assembly, inversion and outcome sorting

The engine assembles Conservative, Central and Expansive scenarios positionally: low values feed the Conservative scenario, central values feed the Central scenario, and high values feed the Expansive scenario. This is reproducible, but it assumes that the input bands move co-monotonically.

Firm-level PDAV = monetised data value / active users

A ratio can invert when the denominator band is proportionally wider than the numerator band. For example, if a firm's active-user estimate expands faster than its revenue estimate, the positional Expansive scenario can produce a lower per-person value than the positional Conservative scenario. This occurs most often for firms with contested user definitions, sparse population evidence, fast user growth or wide regional-share uncertainty.

At display time, per-firm output bands are sorted by outcome: the smallest value is labelled Conservative, the middle value Central and the largest value Expansive. The raw positional scenario values, citations, provenance, confidence scores and input pairings remain intact.

This outcome sorting is defensible because valuation ranges normally describe output values. The caveat is provenance alignment: for an inverted firm, a displayed Conservative value may have originated from positional high inputs. The user-facing label therefore describes the output result, while the stored provenance still describes the input scenario that generated it.

In the current cohort, 27 of the 129 firms exhibit full inversion (~21%) (Conservative > Expansive). A further 28 firms (~22%) exhibit a non-monotonic zig-zag pattern where the Central value falls outside the L–H range. Together, 55 of 129 firms (~43%) have non-monotonic per-firm bands, and the outcome-sort fix is applied to every one of them.

The universe rollup uses positional scenarios before per-firm outcome sorting, so the Conservative cohort line can be biased upward for inverted firms — their positional-L value, which is structurally the largest per-firm contribution, is what enters the cohort Conservative total. The display-time fix re-labels per-firm cards but does not re-compute the cohort rollup. With 27 fully-inverted firms out of 129, the cohort Conservative bias is non-trivial but small relative to the band width; the cohort Central and Expansive lines are essentially unaffected.

13. Limitations and model governance

The model is designed to be defensible, transparent and reproducible. Its principal limitations are parameter uncertainty, category simplification, regional defaulting and the deterministic assembly of scenario bands.

Issue	**Technical implication**
Four archetypes are not exhaustive	The scheme supports parameterisation but may describe hybrid firms only approximately.
ψ^attr defaults are literature-anchored rather than measured for every firm	Firm-specific overrides improve precision where evidence exists; smaller firms may rely on archetype defaults.
AI archetype assignment is an AI-applied rubric (transparent rubric-AI classifier)	The classifier is deterministic over a fixed rubric, so its bias is consistent and re-runnable when the rubric is refined; it is not equivalent to a human expert panel.
Regional shares use defaults for some firms	Default regional revenue or subject shares can understate or overstate regional exposure depending on firm geography.
Paying-customer counts can use fallback shares	Where firms do not disclose paying-customer counts, the model estimates them from active users and archetype-specific paying-user shares.
Scenario bands are sensitivity bands	The Conservative and Expansive bands are not 95 per cent confidence intervals.
The parameter set is opinionated	Alternative methodologists could draw channels, shares or population allocations differently. The schema is designed so those choices can be revisited.
Scenario assembly assumes co-monotonicity	Independent uncertainty is compressed into positional bands; a Monte Carlo implementation would better represent uncertainty.

The strongest uncertainty statement is that the current banding system expresses sensitivity to selected input positions. It does not model a full probability distribution over independent variables. In reality, revenue, user counts, regional shares and parameter overrides have independent uncertainties that may offset or compound.

14. Reproducibility and audit model

Each calculation run is reproducible from the model version, cohort definition, region, scenario band, period year, source inputs (including overlap fields), parameter resolution and provenance tags, and the population cap in force at compute time. The audit model distinguishes four data layers and one run-trail layer:

- **Source layer**: evidence pages, filings, working papers, estimates and analyst notes, attached to every input through citation records and the supporting AI evidence, justification, key findings, confidence score, model identity and timestamp.
- **Input layer**: low, central and high estimates for revenues, populations, overlap fields and parameter overrides, stored with confidence and provenance.
- **Parameter layer**: model defaults, archetype values and firm-specific overrides, resolved at compute time in that precedence order and tagged with a provenance flag (entity, archetype blend, archetype, model default, unknown).
- **Output layer**: persisted firm metrics and universe snapshots, keyed by model, cohort, region, scenario and period year.
- **Run-trail layer**: every recompute is recorded as a calculation-run row capturing the operator, timestamp, scenarios executed, warnings raised and final status, so any historical output value can be traced back to the operator, the inputs that produced it and the moment it entered the audit record.

Every persisted firm-metric row and universe-snapshot row also carries an inputs_hash, a SHA-256 over the canonical-JSON inputs that produced it. A re-run on identical inputs produces an identical hash; a re-run on altered inputs produces a different hash, so input drift is detectable without re-deriving the output values.

Combined with the run-trail layer, any historical headline can be byte-for-byte verified, and any divergence between a re-run and the stored value localises the change in the underlying inputs.

15. References

[1] Aridor, G. and Che, Y.-K. (2024) Privacy Regulation and Targeted Advertising: Evidence from Apple's App Tracking Transparency. Working paper, January 19, 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4698374

[2] Aridor, G., Che, Y.-K., Hollenbeck, B., Kaiser, M. and McCarthy, D. (2025) Evaluating the Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apple's App Tracking Transparency. CESifo Working Paper No. 10928 (Original Version: January 2024; This Version: May 2025). https://www.ifo.de/en/cesifo/publications/2024/working-paper/evaluating-impact-privacy-regulation-e-commerce-firms-evidence

[3] Veldkamp, L. (2023) Valuing Data as an Asset. Review of Finance, 27(5), pp. 1545-1562. https://doi.org/10.1093/rof/rfac073

[4] Tucker, C. (2019) Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility. Review of Industrial Organization, 54, pp. 683-694. https://doi.org/10.1007/s11151-019-09693-7

[5] Kaplan, J. et al. (2020) Scaling Laws for Neural Language Models. arXiv:2001.08361. https://doi.org/10.48550/arXiv.2001.08361

[6] Hoffmann, J. et al. (2022) Training Compute-Optimal Large Language Models. arXiv:2203.15556. https://doi.org/10.48550/arXiv.2203.15556

[7] Ouyang, L. et al. (2022) Training Language Models to Follow Instructions with Human Feedback. Advances in Neural Information Processing Systems 35. https://proceedings.neurips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html

[8] Hong, R., Hutson, J., Agnew, W., Huda, I., Kohno, T. and Morgenstern, J. (2025) A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset. arXiv:2506.17185. https://doi.org/10.48550/arXiv.2506.17185

[9] Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F. and Fox, K. J. (2025) GDP-B: Accounting for the Value of New and Free Goods. American Economic Journal: Macroeconomics, 17(4), pp. 312-344. https://doi.org/10.1257/mac.20210319

[10] World Bank, World Development Indicators. Individuals using the Internet (% of population), sourced from the International Telecommunication Union. https://data.worldbank.org/indicator/IT.NET.USER.ZS

[11] Consensus Economics. Economic Forecasts and Indicators. https://www.consensuseconomics.com/

[12] London Stock Exchange Group. Consensus Economics Data. https://www.lseg.com/en/data-analytics/financial-data/economic-data/international-economic-indicators/business-governance-economic-indicators/consensus-economics-data

[13] International Monetary Fund. World Economic Outlook, April 2026. https://www.imf.org/en/publications/weo/issues/2026/04/14/world-economic-outlook-april-2026

[14] OECD. Inflation forecast indicator. https://www.oecd.org/en/data/indicators/inflation-forecast.html

[15] Board of Governors of the Federal Reserve System. 2025 Statement on Longer-Run Goals and Monetary Policy Strategy. https://www.federalreserve.gov/monetarypolicy/monetary-policy-strategy-tools-and-communications-statement-on-longer-run-goals-monetary-policy-strategy-2025.htm

[16] Bank of Canada. Inflation-control target. https://www.bankofcanada.ca/rates/indicators/key-variables/inflation-control-target/

[17] European Central Bank. Two per cent inflation target. https://www.ecb.europa.eu/mopo/strategy/pricestab/html/index.en.html

[18] Bank of England. Inflation and the 2% target. https://www.bankofengland.co.uk/monetary-policy/inflation

[19] Federal Reserve Bank of St. Louis, FRED. 10-Year Breakeven Inflation Rate. https://fred.stlouisfed.org/series/T10YIE

[20] European Central Bank. Activity and price discovery in euro area inflation-linked swap markets. ECB Economic Bulletin, Issue 5/2025. https://www.ecb.europa.eu/press/economic-bulletin/articles/2025/html/ecb.ebart202505_02~b0c28fc22e.en.html

web3
foundation