Research Question 2: Value-Cost Alignment & Perceived Fairness

To what extent does the pricing structure align the value delivered to different user segments with the costs they incur, and how does this alignment affect adoption, retention, and perceived fairness?

Executive Summary

This research question examines whether per-token pricing models create **equitable value-cost alignment** across diverse user segments, and how perceptions of fairness influence market outcomes. The central tension lies between **cost-based metrics** (tokens processed) and **value-based outcomes** (business objectives achieved). Evidence reveals that token-based pricing, while aligned with provider costs, often **misaligns with customer value** and generates **fairness concerns** that undermine adoption and retention. The analysis synthesizes pricing theory, behavioral economics, and empirical studies to propose value-aligned alternatives.

Theoretical Framework

Value-Based vs. Cost-Based Pricing

Classical pricing theory distinguishes between three fundamental approaches:

1. Cost-Plus Pricing

Price = Cost + Markup. This approach ensures profitability but ignores customer value perception. In digital services, marginal costs approach zero, making cost-plus pricing problematic [1] [2].

2. Competition-Based Pricing

Price set relative to competitors. This approach ignores both costs and value, leading to potential under-pricing (if value exceeds competitors' prices) or over-pricing (if value falls short) [3].

3. Value-Based Pricing

Price reflects the **perceived value delivered to customers**, not the cost to produce [4] [5] [6] [7]. This approach maximizes revenue by capturing consumer surplus while ensuring customers perceive fair exchange.

Token-Based Pricing as Cost-Based Metric

Research on AI pricing patterns reveals that token-based pricing is fundamentally a "**cost-based metric**"—it reflects the **computational resources consumed** by the provider, not the **value delivered** to the customer [8]. As one analysis states: "Token Based. This is generally a cost-based metric and is used by most of the large foundation model companies" [8].

This creates a **value-cost misalignment** when:

- **High-value tasks** consume few tokens (e.g., a single critical code fix)
- **Low-value tasks** consume many tokens (e.g., verbose summaries of trivial content)
- **Task complexity** doesn't correlate with token consumption (a difficult question may yield a brief answer)

Fairness Theory in Pricing

Philosophical and economic theories of fairness provide frameworks for evaluating pricing structures:

1. Procedural vs. Distributive Fairness

- **Procedural fairness**: The process of price-setting is transparent and consistent [9] [10]
- **Distributive fairness**: The outcome (who pays what) aligns with principles of justice [9] [10]

Token-based pricing may achieve procedural fairness (everyone pays the same per-token rate) while violating distributive fairness (outcomes are unequal due to hidden factors like language or task type).

2. Reference Price Theory

Consumers evaluate fairness by comparing actual prices to **reference prices**—what they expect to pay based on:

- **Historical prices** (what they paid before)
- **Competitor prices** (what alternatives cost)
- **Cost estimations** (what they believe it "should" cost) [11] [9]

Research on online platforms demonstrates that when actual prices deviate significantly from reference prices, consumers perceive the pricing as **unfair**, particularly when transparency is lacking [11] [9] [10].

3. Equity Theory

Fairness requires that **outcomes be proportional to inputs**. In pricing contexts:

$$fractextCustomerBenefittextPricePaidapproxtextFairExchangeRatio$$

When this ratio varies significantly across customers for reasons they perceive as arbitrary (e.g., language spoken, query phrasing), fairness concerns arise [11] [12] [13] [14].

Empirical Evidence: Value-Cost Misalignment

The Willingness-to-Pay Gap

Research on AI service pricing reveals a critical finding: the **most common pricing approach** is based on **Willingness to Pay (WTP)**, followed by **Value-Based** pricing, with **Cost-Plus** being uncommon [8]. This suggests that providers recognize the importance of aligning price with customer value.

However, **token-based pricing** contradicts this approach because:

1. **WTP varies by outcome**, not input (tokens consumed)
2. **Value-based pricing** focuses on business impact, not computational cost
3. **Token consumption** is invisible to customers until after delivery

Case Study: AI Content Generation

Consider two users of an AI writing service:

- **User A**: Generates a marketing email (200 tokens) that drives $50,000 in sales
- **User B**: Generates a blog post (2,000 tokens) that receives no engagement

Under token-based pricing:

- User A pays $0.50 (200 tokens × $0.0025/token)
- User B pays $5.00 (2,000 tokens × $0.0025/token)

Under value-based pricing:

- User A should pay MORE (captured significant value)
- User B should pay LESS (captured minimal value)

The misalignment is **10x in the wrong direction**—the user who derived 1000x more value paid 10x less [4] [5] [6].

Customer Stickiness & Value Perception

Research on value-added services in digital platforms demonstrates that **customer stickiness** (loyalty, repeated use) is influenced by:

1. **Perceived value alignment**: Customers who feel pricing reflects the value they receive exhibit higher retention [15]
2. **Consumption patterns**: Sticky customers are less price-sensitive, enabling premium pricing for differentiated value [15]
3. **Service quality perception**: The relationship between price and quality expectations affects satisfaction [15]

For token-based pricing, stickiness is undermined when:

- **Bill shock** occurs (unexpected costs despite perceived low value)
- **Comparison shopping** reveals better value elsewhere (competitors offer outcome-based pricing)
- **Trust erosion** from unpredictable costs (uncertainty breeds disloyalty) [16] [17]

Platform Service Strategies & Purchase Behavior

A study of online platforms optimizing service strategies based on purchase behavior found that platforms with **data-collecting capabilities** should collaborate with sellers to offer services to new consumers, maximizing profits for all parties [18] . This suggests that **optimal pricing strategies** account for:

1. **Behavioral data**: Historical usage patterns predict future value
2. **Segmentation**: Different customer types derive different value from identical services
3. **Dynamic adjustment**: Pricing should evolve as customer relationships mature [18]

Token-based pricing is **static and blind** to these factors—a new user and a loyal customer pay identical per-token rates, regardless of:

- **Learning curve**: New users consume more tokens experimenting
- **Expertise**: Experienced users extract more value per token
- **Integration depth**: Strategic customers derive compounding value over time

This **one-size-fits-all** approach leaves significant value on the table while potentially overcharging low-value users [18] [4] .

Fairness Perceptions in Online Platforms

Transparency & Trust

Research on price fairness in online food service platforms reveals several critical findings:

1. Reference Prices Matter Less Online

Contrary to offline services, **external reference prices** (what competitors charge) showed **insignificant influence** on online price fairness perceptions [9] . Instead, fairness is determined by:

- **Trust** developed through marketing and favorable reviews [9]
- **Perceived quality** (taste, service, ambiance) relative to price paid [9]
- **Alignment** between perceived quality and actual price [9]

Application to Token Pricing: This suggests that even if token prices are **competitively low**, users may perceive unfairness if:

- **Quality signals** are weak (model capabilities poorly communicated)
- **Reviews** highlight bill shock or unexpected costs
- **Marketing** emphasizes low per-token prices but actual bills are high (tokens-per-task problem)

2. Input-Output Relationship Clarity

The study emphasizes that consumers assess fairness based on the **relationship between inputs (price paid) and outputs (quality received)** [9] . In traditional services, this relationship is direct:

- Pay $20 for a meal → Receive a meal of quality X
- If quality X ≥ expected quality for $20 → Perceived as fair

In token-based pricing, the relationship is **mediated and opaque**:

- Pay $20 for N tokens → Receive output of quality X (but N is unknown until after consumption)
- If quality X ≥ expected quality for $20 → Was the token count justified?

This **mediation** creates **fairness uncertainty**—users cannot assess whether the price they paid was fair because they lack understanding of the input-output mapping [19] [20].

Personalized Pricing & Discrimination Concerns

Research on personalized pricing and price discrimination reveals nuanced fairness perceptions:

1. Transparency as Fairness Enabler

When done **transparently**, personalized pricing can be perceived as **fairer** than generic pricing, as it reflects individual circumstances and behaviors rather than arbitrary factors [11]. However, this requires:

- **Clear communication** of why prices differ
- **Justifiable criteria** (e.g., volume discounts, loyalty rewards)
- **User control** over factors affecting their price [11]

2. Fairness Concerns from Opaque Personalization

When personalization is **opaque**, perceived unfairness escalates:

- **97% of respondents** in a global investigation expressed concern about transparency and fairness of personalized pricing [11]
- Customers perceive **exploitation** when they discover others paid less for identical services [11]
- **Algorithmic pricing** can lead to price gouging, exacerbating fairness concerns [3]

Application to Token Pricing: While token pricing appears "fair" (everyone pays the same per-token rate), **hidden personalization** occurs through:

- **Tokenization differences** (language, phrasing affects token count)
- **Model selection** (some models are more token-efficient for certain tasks)
- **System prompts** (provider-controlled instructions consume tokens invisibly)

Users who discover these hidden variations may perceive **deceptive fairness**—a veneer of equality masking structural inequality [21] [22].

Distributive Justice Principles

Philosophical frameworks for distributive justice identify four core principles:

1. Equality: Everyone receives equal shares [12] [13] [14]

Token pricing achieves **formal equality** (same per-token rate) but not **substantive equality** (same cost for same value).

2. Desert: Allocation based on merit or contribution [12] [13] [14]

Token pricing has **no desert component**—users who derive more value don't pay more, and users who contribute more (e.g., through feedback that improves the model) aren't rewarded.

3. Need: Allocation prioritizes those with greater need [12] [13] [14]

Token pricing is **blind to need**—nonprofits, educators, and researchers pay the same as well-resourced enterprises, despite potentially higher social value of their use cases.

4. Efficiency: Maximize total welfare [12] [13] [14]

Token pricing may achieve **allocative efficiency** (resources go to highest-value uses IF users can accurately assess value and costs), but **information asymmetries** undermine this [23] [24] [19].

A **just pricing model** would integrate multiple principles, not rely solely on one. Current token-based approaches **fail** on desert, need, and (due to information failures) efficiency dimensions [12] [13] [14].

Value-Based Pricing Alternatives

Implementation Frameworks

Research on implementing value-based pricing in digital agencies provides actionable frameworks:

1. Benefit Analysis

Conduct **thorough benefit analysis** for each service, identifying:

- **Tangible value**: Revenue increase, cost savings, efficiency gains
- **Intangible value**: Brand enhancement, risk reduction, strategic positioning
- **Quantifiable outcomes**: Metrics that can be measured and attributed [4] [6]

2. Pricing Tiers Based on Value Delivery

Develop tiered structures that reflect **increasing value**, not just increasing consumption:

- **Base tier**: Essential services with foundational outcomes
- **Premium tier**: Enhanced features with demonstrably superior results
- **Enterprise tier**: Strategic partnership with revenue-sharing or outcome guarantees [4] [6]

Each tier should **clearly demonstrate** the incremental value, using:

- **Case studies** showing ROI for similar customers
- **Performance metrics** (e.g., "30% faster processing" not "30% fewer tokens")
- **Value calculators** that translate usage into business outcomes [4]

3. Value Metrics Instead of Token Counts

Rather than exposing raw token consumption, present pricing in **value-aligned metrics**:

- **Per successful outcome**: "$0.50 per correctly answered customer question"
- **Per business function**: "$100/month for unlimited invoice processing"
- **Per user benefit**: "$5 per user per month for email assistance"

This aligns the **unit of payment** with the **unit of value**, making fairness assessment intuitive [20] [25] [4].

Hybrid Models: Balancing Fairness & Predictability

Industry practice reveals that **pure models** (pure subscription, pure usage-based) often fail, and **hybrid approaches** achieve superior outcomes:

1. Tiered Plans with Included Usage

- **Base subscription** provides predictable budget floor
- **Included token allowance** enables typical usage without variability
- **Overage pricing** handles edge cases, priced at premium to incentivize right-tier selection [26] [27]

Benefits:

- **Predictability** for budgeting (subscription component)
- **Fairness** for variable usage (consumption component)
- **Mental accounting** simplification (subscription feels "free," overage feels "extra") [28]

2. Volume Discounts & Committed Use

- **Commit to annual volume** (e.g., 10 million tokens) to receive discount
- **Unused tokens** roll over or convert to credits
- **Excess usage** priced at published rates [26]

Benefits:

- **Aligns with enterprise procurement** (annual budgets)
- **Reduces bill shock** (predictable costs for committed volume)
- **Captures value** from high-volume users (who benefit from discounts) [26]

3. Outcome-Based Pricing

Early experiments show promising results from **tying costs directly to value delivered**:

- **Legal tech**: Casetext offers "completion guarantees"—payment only when document review objectives are achieved

- **Healthcare AI**: Tempus ties pricing to successful diagnostic assistance
- **Customer service**: Level AI bases pricing on measured reduction in resolution times [29]

These models **eliminate value-cost misalignment** by making payment **contingent** on outcome achievement. However, they require:

- **Measurable outcomes** (not all AI value is quantifiable)
- **Attribution clarity** (proving the AI caused the outcome)
- **Risk sharing** (provider absorbs costs if outcomes fail) [29]

Adoption & Retention Implications

The Adoption Paradox

Behavioral economics research reveals a tension in adoption decisions:

1. Initial Appeal of "Fair" Pricing

Usage-based pricing **increases adoption** because:

- **Low commitment**: "Try it, pay only for what you use"
- **Fairness framing**: "You won't pay for what you don't need"
- **Risk reduction**: No sunk cost if the tool doesn't deliver value [16] [17]

2. Post-Adoption Disillusionment

After using the service, retention suffers when:

- **Bill shock** occurs (actual costs exceed expectations) [30] [17] [31]
- **Value misalignment** becomes apparent (paid for tokens, not outcomes)
- **Competitors** offer simpler, more predictable alternatives [32]

This creates **high churn in usage-based models** unless mitigated through:

- **Transparent forecasting** (realistic cost projections)
- **Value communication** (connecting spend to outcomes)
- **Pricing flexibility** (ability to switch to predictable plans) [17]

Lock-In Effects & Switching Costs

Research on lock-in effects in online platforms demonstrates that **reputation and usage data** create switching costs [33] [34] [35]. Applied to AI services:

1. Data Lock-In

- **Fine-tuned models** on proprietary data (switching means losing customization)
- **Historical usage patterns** (new provider lacks context for optimization)

- **Integration depth** (API calls embedded throughout infrastructure) [33]

2. Learning Lock-In

- **Prompt engineering expertise** (users become proficient with specific models)
- **Workflow integration** (business processes built around current provider)
- **Team familiarity** (retraining costs for new tools) [33]

3. Price Lock-In Through Committed Use

- **Volume discounts** require annual commitments
- **Prepaid credits** create sunk cost (pressure to use even if better alternatives emerge)
- **Bundling** with other services (e.g., cloud infrastructure + AI) [26]

Fairness Implications: Lock-in enables **exploitation**—once users are locked in, providers can:

- **Raise prices** (knowing switching costs are high) [33] [32] [34]
- **Reduce quality** (knowing users can't easily leave) [33]
- **Extract surplus** that would otherwise go to customers [33]

Research confirms: "platforms can capitalize on lock-in effects more effectively" when portability is prevented [33] [34]. For token pricing, lock-in manifests through:

- **Proprietary tokenization** (different providers count tokens differently, preventing comparison)
- **Model-specific optimizations** (prompts tuned for GPT-4 may perform poorly on Claude)
- **Vendor-specific tooling** (cost monitoring, optimization tools tied to provider ecosystems)

Policy Response: **Reputation portability** and **data portability** regulations could mitigate lock-in exploitation, improving long-term fairness [33].

Segmentation & Differential Value Capture

User Heterogeneity in Value Derivation

Not all users derive equal value from identical token consumption. Consider:

Segment A: Strategic Enterprise Users

- **Use case**: Automating high-value workflows (contract analysis, code generation)
- **Value per token**: High (each token contributes to major cost savings or revenue)
- **Price sensitivity**: Low (paying for business outcomes, not tokens)
- **Fairness concern**: Undercharged relative to value (willing to pay more) [4]

Segment B: Hobbyist Users

- **Use case**: Personal projects, learning, experimentation
- **Value per token**: Low (primarily entertainment or education value)

- **Price sensitivity**: High (paying out of pocket, limited budget)
- **Fairness concern**: Overcharged if bills are unpredictable (bill shock) [17]

Segment C: Non-English Users

- **Use case**: Identical to English users, but in different language
- **Value per token**: Equivalent to English users (same objective outcomes)
- **Price sensitivity**: Comparable, but **cost is 5-25x higher** due to tokenization
- **Fairness concern**: **Structural discrimination**—paying more for equal value [21] [22]

Optimal Pricing Strategy: **Segment-specific pricing** that reflects value, not just cost:

- **Outcome-based pricing** for enterprises (capture value surplus)
- **Freemium or tiered subscription** for hobbyists (predictable costs)
- **Language-normalized pricing** (same cost for equivalent outputs regardless of tokenization) [4] [21]

Current token-based pricing **fails to segment appropriately**, leaving money on the table with enterprises while potentially overcharging hobbyists and discriminating against non-English users.

Behavioral Pricing & Price Discrimination

Research on behavioral pricing demonstrates that **customer behavior can signal value**:

1. Usage Intensity as Value Signal

High-frequency users typically derive **higher per-unit value** (they wouldn't use the service extensively if value were low). **Volume discounts** capture this:

- **Tiered pricing**: First 1M tokens at $X, next 10M at $X × 0.8, beyond 50M at $X × 0.6
- **Committed use discounts**: Commit to 100M tokens annually, receive 20% discount [26]

This achieves **second-degree price discrimination**—users self-select into tiers based on their value derivation [7] .

2. Feature Access as Value Differentiation

Users who pay for **premium features** (faster response, longer context, advanced models) signal higher value needs:

- **GPT-4 vs GPT-3.5**: Price differential reflects capability gap
- **Priority access**: Pay premium for guaranteed availability during peak times
- **Extended context**: Pay premium for larger context windows [36]

This **feature-based segmentation** aligns payment with value better than raw token counts [7] .

3. Time-Based Differentiation

- **Real-time vs batch processing**: Real-time costs more (higher value use cases)
- **Peak vs off-peak**: Dynamic pricing based on infrastructure load

- **Latency guarantees**: SLAs for response time command premium prices [26]

Comparative Analysis: Token vs. Alternative Models

Token-Based Pricing

Strengths:

- **Theoretically fair**: Pay for resources consumed
- **Aligns with provider costs**: Computational resources directly correlate with tokens
- **Scalable**: Simple to implement and communicate [8]

Weaknesses:

- **Misaligns with customer value**: Value is outcome-based, not token-based
- **Unpredictable costs**: Users can't forecast consumption accurately [24] [19]
- **Information asymmetry**: Provider controls measurement (tokenization) [21] [22]
- **Behavioral challenges**: Mental accounting, bill shock, bounded rationality [28] [30] [17]

Subscription-Based Pricing

Strengths:

- **Predictable**: Fixed monthly cost enables budgeting
- **Mental accounting**: Feels "free" after subscription paid (encourages usage) [28]
- **Simple**: No usage tracking or complex calculations

Weaknesses:

- **Misaligns with usage**: Heavy users subsidized by light users (or vice versa)
- **Inefficient**: Dead weight loss from under-utilization (unused subscriptions)
- **Limited scaling**: Enterprise users hit artificial caps [16]

Outcome-Based Pricing

Strengths:

- **Perfect value alignment**: Pay for results, not inputs
- **Risk sharing**: Provider incentivized to deliver outcomes
- **Fairness**: Only pay if value received [29]

Weaknesses:

- **Attribution challenges**: Hard to prove causation (was outcome due to AI?)
- **Measurement complexity**: Not all outcomes are easily quantified
- **Revenue uncertainty**: Provider can't predict income [29]

Hybrid Tiered Models

Strengths:

- **Balanced**: Predictability (subscription) + fairness (usage component)
- **Segmentation**: Different tiers for different value profiles
- **Flexibility**: Users can adjust tier as needs change [26]

Weaknesses:

- **Complexity**: Harder to communicate than pure models
- **Optimization burden**: Users must select "right" tier [26]

Optimal Approach: Evidence suggests **hybrid tiered models with value-based metrics** achieve the best balance:

- **Base subscription** (predictability)
- **Included usage** measured in **value units** not tokens (e.g., "1000 queries" not "1M tokens")
- **Tiered features** based on business outcomes (e.g., "basic analysis" vs "strategic insights")
- **Outcome guarantees** for enterprise tiers (e.g., "95% accuracy or refund") [4] [26] [29]

Policy & Design Recommendations

Value-Aligned Metric Design

1. Outcome Units Instead of Token Counts

Replace token-based billing with **outcome-based units**:

- **Document processing**: "per document processed" (regardless of length)
- **Customer support**: "per conversation resolved"
- **Content generation**: "per deliverable" (blog post, email, report)
- **Code assistance**: "per successful compilation" or "per test passed"

This makes the **price-value relationship transparent** [20] [25] [4].

2. Capability Tiers Instead of Model Selection

Rather than forcing users to choose models (GPT-4, Claude 3.5, etc.), offer **capability tiers**:

- **Basic**: "Suitable for simple Q&A, drafting"
- **Advanced**: "Handles complex reasoning, long documents"
- **Expert**: "Strategic analysis, multi-step problem solving"

The provider **automatically selects the most cost-efficient model** that meets the tier requirements, optimizing for value delivery rather than token consumption [4] [6].

3. Fairness Metrics & Monitoring

Implement **fairness dashboards** showing:

- **Value delivered per dollar spent** across customer segments
- **Cost consistency** for equivalent outcomes (e.g., same task in different languages)
- **Outcome achievement rates** by tier and use case [12] [13]

Public reporting of fairness metrics creates **accountability** and **competitive pressure** to improve value alignment [11] [9].

Transparency & Justification Requirements

1. Value Explanation

Every invoice should include:

- **What was delivered** (outcomes, not just token counts)
- **How it compares** to expected costs (based on historical usage)
- **Where cost drivers occurred** (which tasks/features consumed most resources) [20]

2. Comparative Context

Provide **benchmarking**:

- "Your cost per outcome is X, which is Y% better/worse than similar users"
- "Alternative approaches would have cost Z"
- "You saved $A by using feature B instead of C" [9]

This enables users to assess **fairness relative to reference points** [11] [9] [10].

3. Contestation Mechanisms

When users perceive unfairness, they should have:

- **Dispute resolution** processes (not just "take it or leave it")
- **Explanation rights** (why did this cost so much?)
- **Adjustment options** (move to different tier, pricing model) [37]

Regulatory & Industry Standards

1. Fairness Certification

Third-party audits assessing:

- **Value-cost alignment** across customer segments
- **Demographic equity** (no systematic disadvantage by language, geography, etc.)
- **Transparency compliance** (adequate disclosure of pricing determinants) [11] [21] [13]

2. Standardized Value Metrics

Industry consortia could develop **benchmark suites**:

- "Standard tasks" (e.g., "summarize 1000-word article")
- "Cost per standard task" as primary comparison metric
- "Value per dollar" ratings across providers [21] [22]

This would enable **true comparison shopping** rather than illusory token-based comparisons that obscure actual costs.

3. Ethical Pricing Guidelines

Building on fairness theory and distributive justice principles:

- Pricing should not **exploit information asymmetries** [23] [37]
- Pricing should **align with value delivered**, not just cost incurred [4] [5]
- Pricing should **not discriminate** based on protected characteristics (language, geography) [21] [22] [12]

Conclusion

The analysis reveals a **fundamental misalignment** between token-based pricing and value-based fairness in AI services:

Theoretical Misalignment: Token pricing is a **cost-based metric** that reflects provider computational costs, not customer value derived. This violates core principles of value-based pricing and creates fairness concerns [4] [8] [6].

Empirical Failures:

- **Willingness-to-pay** research shows value varies by outcome, not token consumption [8] [38]
- **Fairness studies** demonstrate that trust, transparency, and value-alignment drive fairness perceptions—all undermined by token opacity [11] [9] [10]
- **Behavioral research** reveals that usage-based pricing creates anxiety, bill shock, and retention challenges despite initial appeal [30] [17] [31]

Distributional Inequity: Token pricing creates **systematic disadvantages** for:

- **Non-English users** (5-25x higher costs for equivalent value) [21] [22]
- **Novice users** (inefficient token usage due to learning curve)
- **Hobbyists** (disproportionate bill shock relative to budget constraints)

Alternative Approaches: Evidence supports **hybrid models** combining:

- **Value-based metrics** (outcomes, not tokens) [4] [5] [29]
- **Tiered structures** (segmentation by value needs) [26] [6]
- **Outcome guarantees** (payment contingent on results) [29]
- **Fairness monitoring** (transparency and accountability) [11] [12] [13]

The optimal pricing strategy is not **one-size-fits-all** but rather **adaptive to user segments**, **aligned with value delivery**, and **transparent in operation**. Current token-based approaches fail these criteria, suggesting substantial room for innovation in AI services pricing models.

References

Citations are embedded throughout the document using bracketed numbers corresponding to the source IDs from the research phase.

[39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130] [131] [132] [133] [134] [135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [150] [151] [152] [153] [154] [155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165] [166] [167] [168] [169] [170] [171] [172] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211]

⁂

1. http://arxiv.org/pdf/2404.00311.pdf
2. https://www.mdpi.com/2071-1050/14/19/11954/pdf?version=1663838570
3. https://www.abacademies.org/articles/pricing-strategies-in-a-digital-economy-a-microeconomic-perspective-17498.html
4. https://sevenfigureagency.com/implementing-value-based-pricing-in-digital-agencies/
5. https://easydigitaldownloads.com/blog/value-based-pricing-for-digital-products-and-services/
6. https://sevenfigureagency.com/implementing-value-based-pricing-for-digital-agencies/
7. https://www.salesforce.com/sales/cpq/value-based-pricing/
8. https://www.ibbaka.com/ibbaka-market-blog/pricing-patterns-for-generative-ai
9. https://www.sciencedirect.com/science/article/abs/pii/S0278431924003268
10. https://journals.umcs.pl/h/article/download/1742/1357
11. https://verpex.com/blog/marketing-tips/price-discrimination-online-the-fairness-of-personalized-pricing-based-on-user-data
12. https://www.nature.com/articles/s41598-022-19792-3
13. https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2022.883999/full
14. https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13369
15. https://www.mdpi.com/0718-1876/20/3/201
16. https://instituteofinterneteconomics.org/behavioral-economics-of-subscription-pricing/
17. https://flexprice.io/blog/why-ai-companies-have-adopted-usage-based-pricing
18. https://www.mdpi.com/2071-1050/16/19/8545
19. https://www.finops.org/wg/genai-finops-how-token-pricing-really-works/
20. https://kinde.com/learn/billing/billing-for-ai/ai-token-pricing-optimization-dynamic-cost-management-for-llm-powered-saas/
21. https://arxiv.org/pdf/2305.13707.pdf
22. https://aclanthology.org/2023.emnlp-main.614.pdf

23. https://www.tandfonline.com/doi/full/10.1080/01605682.2023.2269212
24. https://www.ikangai.com/the-llm-cost-paradox-how-cheaper-ai-models-are-breaking-budgets/
25. https://www.getmonetizely.com/articles/understanding-token-based-pricing-for-agentic-ai-systems-a-new-paradigm-in-ai-economics
26. https://stripe.com/resources/more/token-pricing-how-it-works-and-how-to-make-the-most-of-it
27. https://techforward.io/the-token-economy-why-usage-based-ai-pricing-is-both-a-blessing-and-a-trap/
28. https://www.behavioraleconomics.com/mental-money-the-psychology-of-subscription-payment-options/
29. https://www.getmonetizely.com/articles/genai-pricing-models-from-tokens-to-outcomes
30. https://stripe.com/resources/more/pricing-flexibility-in-ai-services
31. https://schematichq.com/blog/why-usage-based-billing-is-taking-over-saas
32. https://www.sciencedirect.com/science/article/abs/pii/S1059056022003112
33. https://onlinelibrary.wiley.com/doi/10.1111/jems.12612
34. https://conference.iza.org/DATA_2022/stenzhorn_e32647.pdf
35. https://papers.ssrn.com/sol3/Delivery.cfm/5251923.pdf?abstractid=5251923&mirid=1
36. https://www.solvimon.com/pricing-guides/openai-versus-anthropic
37. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3358724_code2714802.pdf?abstractid=3108333&mirid=1&type=2
38. https://conjointly.com/blog/willingness-to-pay/
39. http://www.emerald.com/jeim/article/36/6/1533-1555/205340
40. https://onlinelibrary.wiley.com/doi/10.1002/mde.4472
41. https://www.frontiersin.org/articles/10.3389/fmars.2025.1601322/full
42. https://arxiv.org/abs/2410.13090
43. https://www.mdpi.com/0718-1876/19/2/61
44. http://ledger.pitt.edu/ojs/ledger/article/download/226/214
45. https://arxiv.org/pdf/2101.06210.pdf
46. https://www.frontiersin.org/articles/10.3389/fphy.2021.631659/pdf
47. https://arxiv.org/pdf/2307.16874.pdf
48. https://arxiv.org/pdf/2208.10271.pdf
49. https://dl.acm.org/doi/pdf/10.1145/3649318
50. https://www.rairo-ro.org/articles/ro/pdf/2022/01/ro210226.pdf
51. https://backend.orbit.dtu.dk/ws/files/290548890/2022_TMG_PartC_1_.pdf
52. https://arxiv.org/pdf/2410.19107.pdf
53. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4437801_code2617082.pdf?abstractid=3613261&mirid=1
54. https://skywork.ai/skypage/en/Technical-Barriers-to-Entry:-Challenges-in-North-American-AI-Model-Localization-and-Implementation/1950070050110914560
55. https://dl.acm.org/doi/fullHtml/10.1145/3497701.3497733
56. https://www.getmonetizely.com/articles/how-to-design-effective-pricing-models-for-network-effects-and-platform-businesses
57. https://competition-bureau.canada.ca/en/how-we-foster-competition/education-and-outreach/consultation-artificial-intelligence-and-competition-what-we-heard

58. https://en.wikipedia.org/wiki/Two-sided_market
59. https://www.sciencedirect.com/science/article/abs/pii/S0304405X23000715
60. https://www.mercatus.org/research/working-papers/data-really-barrier-entry-rethinking-competition-regulation-generative-ai
61. http://arxiv.org/pdf/2502.16363.pdf
62. https://competitionpolicyinternational.com/assets/Uploads/Autumn2014Schmalensee.pdf
63. https://www.nfx.com/post/network-effects-manual
64. https://www.linkedin.com/pulse/global-ai-llm-market-critical-analysis-through-five-uzwyshyn-ph-d--is3dc
65. https://academic.oup.com/jeea/article-pdf/1/4/990/10312916/jeea0990.pdf
66. https://www.tandfonline.com/doi/full/10.1080/13241583.2024.2393933
67. https://ieeexplore.ieee.org/document/10172155/
68. https://www.multiresearchjournal.com/arclist/list-2024.4.6/id-4262
69. https://elibrary.imf.org/openurl?genre=journal&issn=1018-5941&volume=2023&issue=027
70. https://ieeexplore.ieee.org/document/10380440/
71. https://ijsab.com/volume-32-issue-1/6438
72. https://www.mdpi.com/2071-1050/13/17/9762/pdf
73. https://www.frontiersin.org/articles/10.3389/fgwh.2022.696529/full
74. https://jaesj.journals.ekb.eg/article_398940.html
75. https://www.mdpi.com/1911-8074/17/4/133
76. https://fepbl.com/index.php/csitrj/article/view/577
77. https://www.mdpi.com/2071-1050/13/16/8996
78. https://www.mdpi.com/2071-1050/13/16/8996/pdf
79. https://www.mdpi.com/2071-1050/12/1/49
80. https://www.mdpi.com/2079-9292/5/4/65/pdf?version=1475056683
81. https://www.adb.org/sites/default/files/publication/939786/source-multilateral-platform-sustainable-infrastructure.pdf
82. https://www.mdpi.com/1911-8074/17/4/133/pdf?version=1711119244
83. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.821979/pdf
84. https://arxiv.org/pdf/2208.04710.pdf
85. https://www.mdpi.com/2071-1050/12/9/3893/pdf
86. https://www.bci.ca/adapting-risk-models-for-todays-infrastructure-investment-opportunities/
87. https://www.prompts.ai/en/blog/managing-token-level-costs-ai
88. https://ctu.ieee.org/blog/2023/02/03/solutions-to-the-digital-divide-moving-toward-a-more-equitable-future/
89. https://www.pwc.com/gx/en/industries/tmt/digital-infrastructures-defining-moment-on-climate.html
90. https://daijobu.ai/2025/05/19/millions-of-tokens-the-invisible-unit-of-measurement-shaping-modern-ai/
91. https://martinhilbert.net/CheapEnoughWD_Hilbert_pre-print.pdf
92. https://www.ey.com/en_sg/media/podcasts/moneymultiple/2025/06/unlocking-value-in-growing-digital-infrastructure
93. https://www.getmonetizely.com/articles/should-your-ai-agent-use-token-based-or-subscription-pricing

94. https://arxiv.org/pdf/2402.09697.pdf
95. https://premierscience.com/wp-content/uploads/2024/11/pjcs-24-356-.pdf
96. https://www.sciencedirect.com/science/article/pii/S0148296323001157
97. https://www.linkedin.com/pulse/provisioned-capacity-ai-beginners-guide-dedicated-vs-asaf-liveanu-i1mhe
98. https://www.brookings.edu/articles/fixing-the-global-digital-divide-and-digital-access-gap/
99. https://www.semanticscholar.org/paper/cef1b80bd30d8c31beb37bc73cf1f15a37962008
100. https://ieeexplore.ieee.org/document/10825591/
101. https://academic.oup.com/jamiaopen/article/doi/10.1093/jamiaopen/ooaf055/8161131
102. https://arxiv.org/abs/2508.19008
103. https://mededu.jmir.org/2025/1/e67244
104. https://ieeexplore.ieee.org/document/11170906/
105. https://www.mdpi.com/0718-1876/17/4/63/pdf?version=1664164923
106. https://ej-ai.org/index.php/ejai/article/view/82
107. https://www.cureus.com/articles/350635-preparing-for-vascular-surgery-board-certification-a-comparative-study-using-large-language-models
108. https://theaspd.com/index.php/ijes/article/view/10923
109. https://ascopubs.org/doi/10.1200/JCO.2025.43.16_suppl.e21598
110. http://arxiv.org/pdf/2503.18129.pdf
111. https://arxiv.org/pdf/2502.07736.pdf
112. https://arxiv.org/html/2410.17950
113. https://arxiv.org/pdf/2407.10834.pdf
114. https://arxiv.org/pdf/2402.11754.pdf
115. http://arxiv.org/pdf/2406.06565.pdf
116. https://arxiv.org/pdf/2409.01666.pdf
117. https://www.econtribute.de/RePEc/ajk/ajkdps/ECONtribute_258_2023.pdf
118. https://www.aipricingcomparison.com/text-generation-api-pricing-calculator
119. https://www.econstor.eu/bitstream/10419/171619/1/wp-13-176_rev.pdf
120. https://anotherwrapper.com/tools/llm-pricing
121. https://www.hec.ca/finance/Fichier/Mimra2013.pdf
122. https://arxiv.org/abs/2403.06150
123. https://langtail.com/llm-price-comparison
124. https://www.sciencedirect.com/science/article/pii/S0899825623000726
125. https://www.younium.com/blog/usage-based-pricing
126. https://jurnal.unikal.ac.id/index.php/hk/article/view/3664
127. https://arxiv.org/abs/2509.06069
128. https://www.tandfonline.com/doi/full/10.1080/09537325.2022.2088342
129. https://link.springer.com/10.1007/s10479-021-04036-w
130. https://www.tandfonline.com/doi/full/10.1080/07421222.2023.2229122
131. https://www.ssrn.com/abstract=5218518

132. https://breached.company/red-sea-cable-cuts-the-hidden-crisis-threatening-global-internet-infrastructure/
133. https://journals.sagepub.com/doi/10.1177/10591478241305333
134. https://journal.uinjkt.ac.id/index.php/etikonomi/article/view/33892
135. https://www.semanticscholar.org/paper/f61fed43aa07694fa1df0a4ead140ed1ac39a4bf
136. https://www.semanticscholar.org/paper/e7fbad668e5b950d901cf706b8b300b2a28958c6
137. http://arxiv.org/pdf/1007.4586.pdf
138. https://linkinghub.elsevier.com/retrieve/pii/S0148296322005689
139. https://arxiv.org/pdf/1904.05656.pdf
140. http://www.scholink.org/ojs/index.php/ibes/article/download/4410/4994
141. https://arxiv.org/pdf/2303.13295.pdf
142. https://nottingham-repository.worktribe.com/preview/943810/Cred_SubEval_EJ-Style.pdf
143. https://link.springer.com/10.1007/s10257-025-00702-9
144. https://downloads.hindawi.com/journals/jam/2023/4456931.pdf
145. https://aisel.aisnet.org/ecis2018_rp/147/
146. https://journals.sagepub.com/doi/10.1177/20539517211069632
147. https://www.econstor.eu/bitstream/10419/238225/1/2020-01.pdf
148. https://www.quantilope.com/resources/how-to-conduct-pricing-research-using-conjoint-analysis
149. https://pure.mpg.de/pubman/item/item_3031522_6/component/file_3501197/2019_03online.pdf
150. https://www.productfocus.com/willingness-to-pay-the-hidden-engine-behind-effective-pricing/
151. https://arxiv.org/pdf/2105.01441.pdf
152. https://labs.adaline.ai/p/token-burnout-why-ai-costs-are-climbing
153. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.821979/full
154. https://www.sciencedirect.com/science/article/abs/pii/S0377221724006362
155. https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-025-12321-8
156. https://psychotricks.com/bounded-rationality/
157. https://www.tamarly.ai/blog-2-1/melvines-ai-analysis-12-understanding-tokens-and-the-costs-of-large-language-models-llms-for-enterprises
158. https://www.sciencedirect.com/science/article/abs/pii/S0925527321001225
159. https://ijsi.in/wp-content/uploads/2025/07/18.02.024.20251003.pdf
160. https://thedecisionlab.com/biases/bounded-rationality
161. https://cloudwars.com/ai/enterprise-ai-minute/breaking-down-token-based-pricing-for-generative-ai-large-language-models-llms/
162. https://www.investopedia.com/terms/a/asymmetricinformation.asp
163. https://ojs.apspublisher.com/index.php/amit/article/download/391/300/769
164. https://www.renascence.io/journal/bounded-rationality-customers-simplified-decision-making-processes
165. http://www.emerald.com/jeim/article/34/5/1429-1451/216071
166. https://ieeexplore.ieee.org/document/9934060/
167. https://www.ijecs.in/index.php/ijecs/article/view/4447
168. https://noonomy-journal.ru/images/3_1_2024/3_1_10.pdf

169. https://ieeexplore.ieee.org/document/10502110/
170. https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-024-10777-8
171. https://link.springer.com/10.1007/978-3-031-43185-2_10
172. https://journals.sagepub.com/doi/10.1177/10946705231173116
173. https://www.allsocialsciencejournal.com/search?q=SER-2025-3-064&search=search
174. https://www.iiakm.org/ojakm/articles/2023/OJAKM_Volume11_2pp1-24.php
175. http://arxiv.org/pdf/1506.06648.pdf
176. https://journals.sagepub.com/doi/pdf/10.1177/10946705231173116
177. http://arxiv.org/pdf/2503.21448.pdf
178. https://www.mdpi.com/2227-7072/6/4/87/pdf
179. https://www.mdpi.com/2071-1050/13/24/13701/pdf?version=1639475248
180. https://arxiv.org/html/2407.05484v1
181. http://www.tandfonline.com/doi/abs/10.1080/00207543.2014.922707
182. https://socsc.ktu.lt/index.php/Social/article/view/14247/7540
183. https://www.willingnesstopay.com/webinar/agentic-ai-pricing-4-of-6-ai-pricing-models---part-2-tokens-credit-systems
184. https://www.youtube.com/watch?v=ZHIwPwAPzlA
185. http://mecs-press.org/ijieeb/ijieeb-v15-n3/v15n3-3.html
186. https://ges.jvolsu.com/index.php/en/component/attachments/download/1848
187. https://www.semanticscholar.org/paper/974665e62c139c842cb12359ea08e20222904f10
188. https://www.semanticscholar.org/paper/2875e7d26ede8bbc59a0d4bc2e187d369e72a15c
189. https://www.semanticscholar.org/paper/62d07091dfd8deb3f688b7599e563e04534ab415
190. https://journals.sagepub.com/doi/10.1016/j.ausmj.2019.07.002
191. https://www.emerald.com/insight/content/doi/10.1108/JIDE-08-2021-0004/full/pdf?title=the-achilles-tendon-of-dynamic-pricing-the-effect-of-consumers-fairness-preferences-on-platforms-dynamic-pricing-strategies
192. https://www.ccsenet.org/journal/index.php/ibr/article/download/66540/36058
193. https://arxiv.org/pdf/2311.00846.pdf
194. https://www.mdpi.com/0718-1876/18/3/60/pdf?version=1688609520
195. https://ejbe.org/EJBE2021Vol14No28p107-J-GOTMARE.pdf
196. https://asistdl.onlinelibrary.wiley.com/doi/10.1002/pra2.2015.145052010043
197. https://journals.sagepub.com/doi/10.1177/21582440241293304
198. https://www.tandfonline.com/doi/pdf/10.1080/1331677X.2020.1844587?needAccess=true
199. https://pmc.ncbi.nlm.nih.gov/articles/PMC10361766/
200. https://www.fiegenbaum.solutions/en/blog/dramatic-drop-ai-token-prices-opportunities-challenges-sustainability
201. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1162916/full
202. https://journals.sagepub.com/doi/10.3233/ISU-240230
203. https://www.untaylored.com/post/demystifying-the-lock-in-business-model-a-comprehensive-explanation
204. https://bear.warrington.ufl.edu/brenner/mar7588/Papers/thaler-mktsci1985.pdf
205. https://www.sciencedirect.com/science/article/pii/S305070062500043X

206. https://www.ewadirect.com/proceedings/aemps/article/view/25070
207. https://pubsonline.informs.org/doi/10.1287/mnsc.2023.4917
208. https://onlinelibrary.wiley.com/doi/10.1002/mde.4460
209. https://pubsonline.informs.org/doi/10.1287/mnsc.2022.4530
210. https://s-lib.com/en/issues/eiu_2025_05_v8_a22/
211. https://www.mdpi.com/0718-1876/20/4/286