

AI Tokens & Pricing

Genesys PureCloud AI Tokens & Pricing Documentation

Study Notes

Topic	Description
Token Model	Consumption-based pricing for AI features
Monthly Allocation	250 tokens for named users, 350 for concurrent users
Pricing	~\$1.00 per additional token beyond allocation
Consumption Tracking	Per-feature metering with detailed reporting
Fair Use Policy	Most customers covered by included allocation
Billing	Monthly renewal with no carryover

Navigation

Admin → Billing & Subscriptions → AI Experience Tokens OR Reports → Usage → AI Token Consumption

AI Tokens Overview

Genesys Cloud implements a comprehensive token-based pricing model for AI features across CX 1-4 license tiers. The platform offers a suite of AI capabilities including Agent Copilot, AI Scoring, Translation, Virtual Agent services, predictive routing, and analytics with flexible consumption-based pricing.

Genesys Cloud AI Experience tokens help you monitor and manage feature consumption and offer flexibility for changing business needs. Tokenization in AI is a way to track AI engagement in real time by allocating fixed units of measurement to usage costs. This allows organizations to pay only for what they use, providing a scalable, cost-efficient way to integrate AI into operations.

Default Token Allocations

- **Named User Model:** 250 tokens per user per month
- **Concurrent User Model:** 350 tokens per month
- **Organization Level:** Shared token pool across all users
- **Renewal:** Monthly renewal with no carryover
- **Additional Tokens:** Available at ~\$1.00 per token (pricing varies by currency)

Edition & Module Requirements

Requirement	Details
Minimum Edition	All Genesys Cloud CX tiers include base tokens
Token Access	All CX 1-4 organizations receive default allocation
Ordering	New customers have access by default
Additional Tokens	Purchase through AppFoundry or Genesys representative
Billing	Included on monthly invoice with consumption tracking

Study Notes - Token Consumption by Feature

Feature	Unit	Tokens Required	Use Case
Agent Copilot	Per user	40-60/month	Real-time agent assistance
AI Scoring	Per evaluation	1 token per 20 evals	Quality management

Feature	Unit	Tokens Required	Use Case
AI Translate	Per minute	1 token per 100 min	Multilingual transcripts
Voice Bots	Per minute	1 token per 17 min	IVR automation
Digital Bots	Per session	1 token per 51 sess	Chat/messaging automation
Messaging	Per message	1 token per 400 msg	Direct messaging channels
Social	Per post	1 token per 400 post	Social media listening
Virtual Agent	Per session	2 tokens per session	Autonomous conversation
Predictive Routing	Per interaction	Included	Agent routing optimization
Analytics	Per user	30-45/month	Analytics access

Token Consumption Details

Agent Copilot

- **Consumption:** 40-60 tokens per user per month
- **Factors Affecting:**
 - Interaction volume (more interactions = more tokens)
 - Knowledge article access frequency
 - AI feature usage intensity (summaries, wrap-up code suggestions)
 - Number of agents using Copilot
- **Example:** 100 agents with heavy Agent Copilot use = 4,000-6,000 tokens/month
- **Recommendation:** Budget 50 tokens per user as safe estimate

AI Scoring (Virtual Supervisor)

- **Consumption:** 1 token per 20 evaluations
- **Calculation:**
 - $100 \text{ interactions/day} \times 20 \text{ working days} = 2,000/\text{month}$
 - $2,000 \div 20 = 100 \text{ tokens/month}$
- **Example:** Medium contact center (400 interactions/day) \approx 400 tokens/month
- **Optimization:** Batch evaluations when possible

AI Translate

- **Consumption:** 1 token per 100 minutes of transcript
- **Use Case:** Converting multilingual transcripts
- **Example:** 10 hours of multilingual review/month = ~6 tokens
- **Application:** Quality management, supervisor reviews

Voice Bots

- **Consumption:** 1 token per 17 minutes
- **Calculation:**
 - 1,000 bot interactions × 5 min avg = 5,000 minutes
 - $5,000 \div 17 = \sim 294$ tokens/month
- **Example:** Heavy bot deployment = 300-500 tokens/month
- **Optimization:** Improve bot deflection rate to reduce minutes

Digital Bots

- **Consumption:** 1 token per 51 sessions
- **Calculation:**
 - 1,000 bot sessions/month = $1,000 \div 51 = \sim 20$ tokens
- **Example:** Moderate chat bot use = 50-100 tokens/month
- **Optimization:** End sessions efficiently to reduce token use

Messaging Channels

- **Consumption:** 1 token per 400 messages
- **Channels:** Facebook, Instagram, WhatsApp, X (Twitter), Apple Business Chat
- **Calculation:**
 - 10,000 messages/month = $10,000 \div 400 = 25$ tokens
- **Example:** Multi-channel deployment = 100-500 tokens/month depending on volume
- **Optimization:** High-volume channels provide better token efficiency

Social Media Listening & Responses

- **Consumption:** 1 token per 400 posts
- **Application:** Social listening and outbound responses
- **Calculation:**
 - 2,000 social posts/month = $2,000 \div 400 = 5$ tokens
- **Example:** Active social presence = 50-200 tokens/month
- **Optimization:** Focus on highest-value channels

Virtual Agent Sessions

- **Consumption:** 2 tokens per session
- **Session Definition:** Single customer interaction on any channel
- **Calculation:**
 - 100 virtual agent sessions/day × 20 days = 2,000 sessions
 - 2,000 × 2 = 4,000 tokens/month
- **Example:** High-volume automation = 2,000-5,000+ tokens/month
- **Optimization:** Improve automation rates to use fewer sessions

Predictive Routing

- **Consumption:** Included (no token cost)
- **Benefit:** Available at no additional cost
- **Note:** Foundational AI feature included with platform

Speech & Text Analytics

- **Consumption:** 30-45 tokens per user per month
- **Factors:** Licensing type (supervisory vs. analyst)
- **Typical Range:** 500-2,000 tokens/month depending on user count

Predictive Engagement

- **Consumption:** No token charges
- **Note:** Included AI capability without token consumption

Token Allocation Examples

Small Contact Center (50 agents)

Configuration:

├ Agent Copilot: 50 agents × 50 tokens = 2,500

├ AI Scoring: ~200 interactions/day

| └ 200 × 20 days ÷ 20 = 200 tokens

└ Messaging: ~500 messages/month

| └ $500 \div 400 = 2$ tokens

└ Virtual Agent: ~50 sessions/day

└ $50 \times 20 \times 2 = 2,000$ tokens

Total Monthly: ~4,700 tokens

Included: $250 \times 50 = 12,500$ tokens

Overage: None (well under allocation)

Cost: \$0 overage

Mid-Market Center (200 agents)

Configuration:

└ Agent Copilot: 200 agents \times 55 tokens = 11,000

└ AI Scoring: ~500 interactions/day

| └ $500 \times 20 \div 20 = 500$ tokens

└ AI Translate: ~5 hours/month

| └ $300 \text{ min} \div 100 = 3$ tokens

└ Messaging: ~5,000 messages/month

| └ $5,000 \div 400 = 13$ tokens

└ Voice Bots: ~1,000 interactions \times 5 min

| └ $5,000 \div 17 = 294$ tokens

└ Virtual Agent: ~200 sessions/day

└ $200 \times 20 \times 2 = 8,000$ tokens

Total Monthly: ~19,810 tokens

Included: $250 \times 200 = 50,000$ tokens

Overage: None

Cost: \$0 overage

Enterprise Center (1,000 agents)

Configuration:

└ Agent Copilot: 1,000 agents \times 50 tokens = 50,000

└ AI Scoring: ~2,000 interactions/day

| └ $2,000 \times 20 \div 20 = 2,000$ tokens

└ AI Translate: ~50 hours/month

| └ $3,000 \text{ min} \div 100 = 30$ tokens

- └ Messaging: ~50,000 messages/month
 - | └ 50,000 ÷ 400 = 125 tokens
- └ Voice Bots: ~5,000 interactions × 4 min
 - | └ 20,000 ÷ 17 = 1,176 tokens
- └ Digital Bots: ~10,000 sessions/month
 - | └ 10,000 ÷ 51 = 196 tokens
- └ Virtual Agent: ~1,000 sessions/day
 - | └ 1,000 × 20 × 2 = 40,000 tokens

Total Monthly: ~93,527 tokens

Included: 250 × 1,000 = 250,000 tokens

Overage: None

Cost: \$0 overage

Note: This enterprise center is well within allocation even with heavy AI usage across all capabilities

Token Consumption Monitoring

Token Usage Dashboard

Organization Token Summary

Period: March 2026

Total Allocation: 50,000 tokens/month

Total Consumed: 19,810 tokens

Remaining: 30,190 tokens (60% available)

Usage Rate: 40%

Consumption by Feature:

- └ Agent Copilot: 11,000 tokens (55%)
- └ Virtual Agent: 8,000 tokens (40%)
- └ AI Scoring: 500 tokens (3%)
- └ Voice Bots: 294 tokens (1.5%)
- └ Messaging: 13 tokens (0.1%)
- └ AI Translate: 3 tokens (0.1%)

└ Other: 0 tokens

Consumption Trends:

└ Week 1: 4,500 tokens (15% of monthly)

└ Week 2: 4,200 tokens (14% of monthly)

└ Week 3: 5,300 tokens (18% of monthly)

└ Week 4: 5,810 tokens (19% of monthly)

└ Trend: Stable, slight increase toward month-end

Projected Usage (if trend continues):

└ End of Month: 19,810 tokens (40% of allocation)

Health Status: ✓ GREEN (well within limits)

Recommendation: Can safely increase AI usage

How to Access Token Monitoring

1. Log into Genesys Cloud Admin
2. Navigate to Reports → Usage
3. Select "AI Token Consumption"
4. Choose time period to review
5. View consumption by feature
6. Export data if needed

Key Metrics to Track

Metric	Purpose	Good Range
Overall Utilization	Total token usage	40-80% of allocation
Per-Feature Usage	Identify heavy consumers	Proportional to use
Trend Direction	Usage pattern	Stable or controlled
Overage Risk	Approaching limits	Monitor if >80%
Unused Allocation	Wasted capacity	<20% unused

Cost Management Strategies

Optimization Techniques

- **Improve Bot Deflection** - Reduce voice bot minutes through better automation
- **End Sessions Properly** - Digital bot sessions end cleanly to reduce count
- **Optimize Routing** - Use Predictive Routing (included) instead of manual methods
- **Batch Processing** - Group analytics and translations for efficiency
- **Selective Features** - Enable only needed AI features per queue
- **Clean Knowledge Base** - Reduce unnecessary knowledge article access

Budgeting Approach

1. **Baseline Calculation:**
 - 50 tokens per Agent Copilot user/month
 - 2 tokens per Virtual Agent session
 - 1 token per 20 evaluations
 - 1 token per 51 digital bot sessions
2. **Add Buffer:**
 - For unpredictability: +20% overage buffer
 - For growth: +10-15% growth buffer
3. **Monitor Monthly:**
 - Review actual vs. projected
 - Identify variance
 - Adjust next month's budget
4. **Planning:**
 - Budget for peak season
 - Factor in new initiatives
 - Plan for scaling

Real-World Budget Example

Mid-Market Center - Annual Budget Planning:

Current State (March 2026):

└ Monthly consumption: 19,810 tokens

└ Monthly allocation: 50,000 tokens

└ Utilization: 40%

└ Cost: \$0 overage

Planned Growth (Next 12 months):

└ +50 agents (10% growth)

- └ New Virtual Agent deployment
- └ Expanded Messaging channels
- └ Expected consumption increase: +8,000 tokens/month

Future State Projection:

- └ Monthly consumption: ~27,800 tokens
- └ Monthly allocation: 62,500 tokens (250 × 250 users)
- └ Utilization: 44%
- └ Cost: \$0 overage

Annual Cost Impact:

- └ Base licensing: No increase (same CX tier)
- └ AI Tokens: No additional cost (within allocation)
- └ Total Additional Cost: \$0 (internal reallocation only)

Conclusion: Growth sustainable within existing token allocation

Token Pricing by Region

Pricing Variations

- **North America:** ~\$1.00 per token
- **Europe:** Varies by currency/region
- **APAC:** Pricing adjusted for region
- **Other Regions:** Contact Genesys for pricing

Ordering Additional Tokens

- **Option 1:** AppFoundry - Self-service purchasing
- **Option 2:** Genesys Representative - Volume discounts available
- **Payment:** Monthly billing or prepaid options
- **Minimum:** No minimum purchase requirement
- **Flexibility:** Increase or decrease allocation monthly

Real-World Usage Scenarios

Scenario 1: Seasonal Spike Management

Contact Center Business: Holiday Retailer

Challenge: Heavy AI usage Dec-Jan, minimal Jun-Aug

Strategy:

- └ Base allocation: 30,000 tokens/month
- └ November: Monitor usage, prepare for peak
- └ Dec-Jan: Purchase additional 20,000 tokens
 - └ Ensures sufficient capacity
 - └ Cost: ~\$20,000 for 2 months
- └ Feb: Return to base allocation
- └ Jun-Aug: Only use allocation (excess unused)

Cost Management:

- └ Off-peak: \$0 additional cost
- └ Peak months: \$20,000 additional
- └ Annual cost: \$20,000 (2 months only)
- └ Flexibility: Scale as needed

Scenario 2: Phased AI Rollout

Contact Center: Implementing AI progressively

Phase 1 (Month 1): Agent Copilot only

- └ Consumption: 4,000 tokens (10 users)
- └ Cost: \$0 (within allocation)
- └ ROI: Proven before major investment

Phase 2 (Month 3): Add Virtual Agent

- └ New consumption: +3,000 tokens
- └ Total: 7,000 tokens
- └ Cost: \$0 (still within allocation)
- └ Result: Expand automation safely

Phase 3 (Month 6): Add Analytics + Bots

└ New consumption: +5,000 tokens

└ Total: 12,000 tokens

└ Cost: \$0 (allocated)

└ Result: Full AI platform

Phase 4 (Month 12): Scale based on ROI

└ Purchase additional tokens if needed

└ Cost: Only what you use

└ Confidence: Proven ROI before major spend

Interview Cheat Sheet

Question	Answer
What is token model?	Consumption-based pricing for AI features
What's the default allocation?	250 tokens per named user, 350 per concurrent user
What's token cost?	~\$1.00 per additional token beyond allocation
How is consumption tracked?	Real-time metering per feature with monthly reporting
Which features consume tokens?	Agent Copilot, Virtual Agent, Bots, Translate, Analytics, etc.
Agent Copilot consumption?	40-60 tokens per user per month
Virtual Agent cost?	2 tokens per session
Voice Bot cost?	1 token per 17 minutes
Digital Bot cost?	1 token per 51 sessions
AI Scoring cost?	1 token per 20 evaluations
Does Predictive Routing cost?	No, included with platform
How do I monitor usage?	Admin → Reports → Usage → AI Token Consumption
Can I purchase additional?	Yes, through AppFoundry or Genesys rep
How often do tokens renew?	Monthly, with no carryover
Is there overage protection?	Fair use policy covers most organizations
What about fair use?	95% of customers covered by fair use policy
How to optimize costs?	Improve bot deflection, optimize routing, batch processing

Question	Answer
Can I predict usage?	Yes, based on user count and AI feature deployment

Key Takeaways

- **Flexible Pricing** - Pay only for what you use
- **Default Allocation** - All CX tiers include base tokens
- **Transparent Metering** - Each feature has clear consumption rates
- **Cost Control** - Budget and monitor usage in real-time
- **Scalability** - Add tokens as needs grow
- **Fair Use** - 95% of customers within standard allocation
- **No Overage Risk** - Purchase additional tokens as needed
- **Monthly Billing** - Consumption tracked and billed monthly
- **Optimization Available** - Multiple strategies to reduce costs
- **Regional Flexibility** - Pricing adjusted by region

Cost Calculation Tool

Quick Estimation Template

Contact Center Token Cost Calculator:

1. Agent Copilot Users: ___ × 50 tokens = ___ tokens

2. Virtual Agent Sessions/Day: ___

Daily calculation: ___ sessions × 2 tokens = ___ tokens/day

Monthly: ___ tokens/day × 20 days = ___ tokens

3. AI Scoring Interactions/Day: ___

Monthly: (___ × 20 ÷ 20) = ___ tokens

4. Voice Bot Minutes/Day: ___

Monthly: (___ × 20 ÷ 17) = ___ tokens

5. Digital Bot Sessions/Month: ___

Monthly: (___ ÷ 51) = ___ tokens

6. Messaging/Month: ___ messages

Monthly: (___ ÷ 400) = ___ tokens

7. Other Features (translate, analytics, etc): ___ tokens

TOTAL MONTHLY CONSUMPTION: ___ tokens

Your Allocation: $250 \times (\text{number of users}) =$ ___ tokens

Overage: (Total - Allocation) or \$0 if under

Annual Cost:

└ Base: Included in licensing

└ Overage: (Monthly overage \times 12) \times \$1.00

└ Total: ___ annually

Additional Resources

Official Documentation

- Token-Based Pricing: help.genesys.cloud/articles/genesys-cloud-tokens-based-pricing-model/
- AI Token Billing: help.genesys.cloud/articles/ai-token-billing/
- AppFoundry Tokens: appfoundry.genesys.com (search "AI Experience Tokens")
- Pricing Overview: genesys.com/pricing

Support & Ordering

- Genesys Sales: sales@genesys.com
 - Genesys Support: <https://support.genesys.com>
 - AppFoundry: <https://appfoundry.genesys.com>
 - Community Forums: <https://community.genesys.com>
-

Document Version Info

Last Updated: March 2026

Source: Genesys PureCloud Official Documentation

Validated: Current with January-March 2026 releases

Version: 1.0

Revision #1

Created 13 March 2026 19:30:56 by Cesar Gzz

Updated 14 March 2026 19:35:03 by Cesar Gzz