

CarClinch Dealership Assistant - Options Analysis

Communication Services (Email/SMS)		
	Azure Communications Service	Twilio
Pricing model	Pay-as-you-go	Pay-as-you-go & Plans (SendGrid for emails)
Projected costs (USD)	<p>~\$0.00025 per email ~\$0.0075 per SMS \$2 /month for business phone number</p>	<p>SendGrid: \$19.95 for < 100,000 emails per month \$89.95 for >100,000 emails per month \$0.0083 per SMS</p>
Pricing notes	Student credit	<p>Free tier available Up to 100 emails/day</p>
SDKs	JavaScript, .NET, Python, Java, iOS, Android, Go (ARM)	Javascript (React), Android, iOS, Python .NET
Security	<p>Encrypted over TLS/HTTPS and at rest Shared responsibility for user access & data Monitoring: Azure Monitor, Activity Log IAM: Microsoft Entra ID Additional security info</p>	<p>Encrypted over TLS/HTTPS and at rest Monitor activity & logs with Twilio Console IAM: OAuth 2.0, API keys, SSO, Twilio Connect</p>
Compliance	Azure Policy	10DLC SMS Guidelines
Scalability	<p>Pay per email & SMS 30 EPM; ACS imposes initial rate limits on emails & must submit support req. to increase rate limits Short code ->100 MPS Toll-Free -> ~4 MPS</p>	<p>Messaging Service to handle growing sender pool SendGrid is fixed monthly plan Pay per SMS Short code -> 100+ MPS Toll-Free -> 3 MPS</p>

CarClinch Dealership Assistant - Options Analysis

AI/LLM Services				
1 Token = ~ 4 English Characters or 0.75 of an english word				
	Azure AI Foundry	OpenAI (GPT)	Claude	Gemini
LLM recommendation	GPT-4o mini GPT-4.1 mini	GPT-4o mini GPT-4.1 mini	Haiku 4.5 Haiku 3.5	Gemini 3 Flash (Preview)
Pricing model	Pay-per-1mil-tokens	Pay-per-1mil-tokens	Pay-per-1mil-tokens	Pay-per-1mil-tokens
Projected costs (USD) (per 1m tokens)	Low 4o mini Input: \$0.15 Output: \$0.60 4.1 mini Input: \$0.40 Output: \$1.60	Low 4o mini Input: \$0.15 Output: \$0.60 4.1 mini Input: \$0.40 Output: \$1.60	High Haiku 3.5 Input: \$0.80 Output: \$4.00 Haiku 4.5 Input: \$1.00 Output: \$5.00	Med-high Input: \$0.50 Output: \$3.00
Pricing notes	Student credits			Free tier/preview Must allow Google to train on input data
Security	Benefits of Microsoft ecosystem. No context training. Standard compliance adherences.	No context training. Standard compliance adherences.	No context training. Standard compliance adherences.	No context training. Standard compliance adherences.

CarClinch Dealership Assistant - Options Analysis

AI/LLM Services				
1 Token = ~ 4 English Characters or 0.75 of an english word				
SDKs	Python, .NET, JavaScript, TypeScript, Java	Python, TypeScript, JavaScript, .NET	Python, TypeScript	Python, JavaScript, Go, Java, .NET
Other notes	<ul style="list-style-type: none">- Ideal if already very integrated with Azure/Microsoft ecosystem- Consider vendor-lockin- Cached results have 1024 token min	<ul style="list-style-type: none">- Lead developer has previous experience working directly with OpenAI API; otherwise, same costs as models in AI Foundry- Mini models are good for general usage- Cached results have 1024 token min	<p>Much better than Gemini/GPT, but robustness and quality of speech/reasoning may be out of scope compared to cost & use case</p> <ul style="list-style-type: none">- Cached results have 2048 and 4096 min b/c Haiku is so fast & relatively cheap	<ul style="list-style-type: none">- Most valuable if there is additional interest in GCP- Cached results have 1024 token min

 Options Analysis - Costs Visualizer