LMasaService.io

Reliable, Responsible, and Cost-Effective Al for Applications and Enterprises

Al Infrastructure for Applications

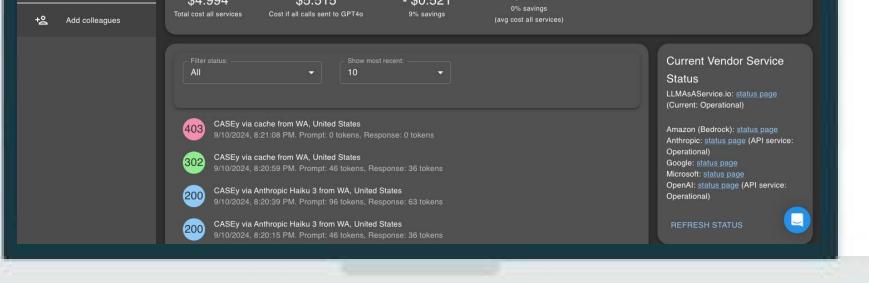
With LLMasaService.io, you can quickly and confidently integrate advanced AI capabilities into your products, empowering your users with cutting-edge functionality. Our platform handles the underlying AI infrastructure, so you can focus on building innovative applications that leverage the power of LLMs.

Managed LLMs for Enterprises

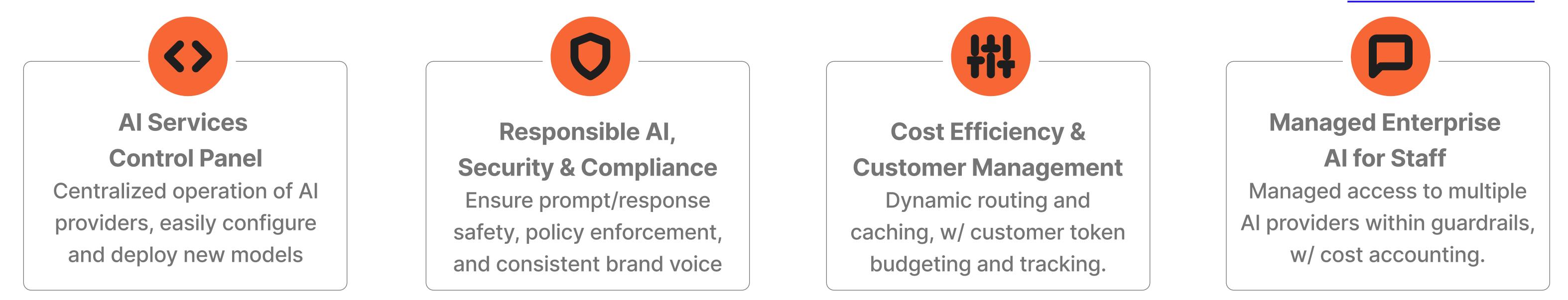
LLMasaService.io's Managed LLM Gateway provides enterprises with an easy-to-use control panel, dynamic routing, policy guardrails,

🞇 LLM as a Service	- add LLM features to your app			
Shark-Puddle (original) -	Control panel for: Shark-Puddle (orig Configuration for Shark-Puddle.com app	inal)		
Trial Tokens Remaining: -43328	Load Balancing Caching M	Customers Load Shedding	Maintenance Mode	Master Switch
A Home	failover method 1800 seconds	Responding to all customers		OFF ON (allow traffic) Status: active
↓↓ LLM Services ↓↓ Route	PROJECT SETTINGS			
Customers	Coat broakdown far laat 20 dawa			
Prompt Library	Cost breakdown for last 30 days	One of Anthropic		4°-41
Embed / Test	\$3.429 \$0.089	Google Anthropic \$0.431 \$0.991	\$0.055 \$	1istral 0.000
 Quality / Safety Enterprise LLM 	Total cost for OpenAl (estimated)	Total cost for Google Total cost for Anthropic	Total cost for Amazon Total c	ost for Mistral
لمن المعامة Integration API	Total Cost Reference Cost \$4 994 \$5 515	Routing Saved Caching S		

off-topic filters, and cost accounting features. This allows you to empower your employees with the superpowers of AI, while maintaining control over usage, costs, and compliance.



More info at Ilmasaservice.io



Benefits of LLMasaService.io:





2

3

Operate Multiple LLMs and Providers with Ease

- Automatic failover between multiple AI providers ensures uninterrupted service.
- Intelligent routing to faster/cheaper/specialized models
- Easily configure, test, and onboard new models/providers

Restrict Confidential And Sensitive Information

- Pll and Sensitive Information can be blocked and redacted before being sent to any LLM vendor.
- Eliminate toxic and unsafe prompts and responses.
- Full call logging and auditing of prompts and responses

Ensure Cost Efficiency and Allocation

- Reduce costs by sending simple prompts to faster, cheaper models and caching repeated requests.
- Allocate costs to customers/departments
- Avoid vendor lock-in and easily bring new models online

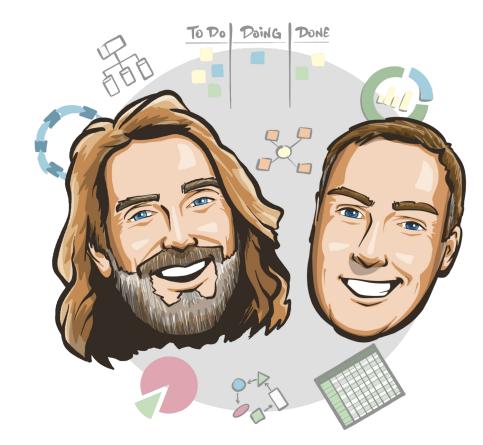
Application Developers

- Add AI features to SaaS, mobile, or web applications without heavy infrastructure investment.
- All applications will eventually include Al features -
- they can code their own infrastructure or use ours. **Enterprises**
- Allow staff managed access to Al for day-to-day use and building into internal applications
- All enterprises will have to govern internal use of Al, and our solution controls cost & avoids vendor lock-in

The Team

Chris Hefley (chris@llmasaservice.io)

- cofounder-CEO of LeanKit, acquired
 - by Thoma Bravo in 2017





Deploy Quality AI features with Confidence

- Unified developer API for LLM interactions
- Streaming, Multi-Turn conversations and UI Components
- Easily update prompts and model routing w/o code changes

Troy Magennis (troy@llmasaservice.io)

Agile product development and

forecasting/metrics expert

focusedobjective.com

Troy and Chris also collaborated on LeanKit from 2012-2017

Want to see it in action?



Check out shark-puddle.com

for a fun app built using

LLMasaService.io

