

オンプレ  
環境で!

# 検索拡張生成でLLMの精度向上 RAG搭載システム

LLM(大規模言語モデル)の回答精度を上げるRAG(検索拡張生成)を搭載したシステムをご用意。  
4U、2U、ワークステーションを揃え、さらにGPUもNVIDIAとAMDでお選びいただけます。

## NVIDIA GPU

## 広がる選択肢

## AMD GPU



NVIDIAハイエンド

H100  
搭載

**VT64 Server XS2-4S-8G**  
ペアボーン: SYS-421GE-TNRT3  
筐体: 4Uラックマウント  
CPU: Intel Xeon Silver 4509Y  
(8C/2.6Ghz) ×2  
Mem: 256GB (16GB DDR5-4800  
ECC REG DIMM ×16)  
SSD: 960GB SATA ×2 ※RAID1  
GPU: NVIDIA H100 NVL 94GB ×1

システム保守  
3年オンサイト保守付



AMDハイエンド

MI 300  
搭載

**VT64 Server AI4-4S-4oG**  
ペアボーン: AS-4145GH-TNMR  
筐体: 4Uラックマウント  
APU: AMD Instinct MI300A ×4  
SSD: 960GB NVMe ×2 ※RAID1  
NIC: HDR 100 ×2Port  
IPMI: 1Port

システム保守  
3年オンサイト保守付



**VT64 Server XS2-4S-8G**  
ペアボーン: SYS-421GE-TNRT3  
筐体: 4Uラックマウント  
CPU: Intel Xeon Silver 4509Y  
(8C/2.6Ghz) ×2  
Mem: 2256GB (16GB DDR5-4800  
ECC REG DIMM ×16)  
SSD: 960GB SATA ×2 ※RAID1  
HDD: 4TB SATA ×1  
GPU: NVIDIA Ada L40S 48GB ×1



**VT64 Workstation XS2-4S-Gen4**  
ペアボーン: SYS-741GE-TNRT  
筐体: タワー型  
CPU: Intel Xeon Silver 4509Y  
(8C/2.6Ghz) ×2  
Mem: 256GB (16GB DDR5-4800  
ECC REG DIMM ×16)  
SSD: 960GB SATA ×2 ※RAID1  
HDD: 4TB SATA ×1  
GPU: NVIDIA Ada L40S 48GB ×1



**VT64 Server AE2-4S-8G**  
ペアボーン: AS-4125GS-TNRT2  
筐体: 4Uラックマウント  
CPU: AMD Genoa 9124  
(16C/3.0Ghz) ×2  
Mem: 256GB (16GB DDR5-4800  
ECC REG DIMM ×16)  
SSD: 960GB SATA ×2 ※RAID1  
HDD: 4TB SATA ×1  
GPU: AMD Instinct MI210 64GB ×1



**VT64 Workstation XS2-4S**  
ペアボーン: SYS-740GP-TNRT  
筐体: タワー型  
CPU: Intel Xeon Gold 6526Y  
(28C/2.0Ghz) ×2  
Mem: 256GB (16GB DDR5-4800  
ECC REG DIMM ×16)  
SSD: 960GB SATA ×2 ※RAID1  
HDD: 4TB SATA ×1  
GPU: AMD Instinct MI210 64GB ×1

システムの詳細はお問い合わせください

## サーバに搭載 blueqat RAG

blueqat社  
オリジナル

- わかりやすく、使いやすい設計
- 最新のさまざまなカスタマイズが可能
- クラウドでの導入もご案内できます、ご相談ください



(画面例)

こんなお悩みありませんか？

ローカルデータを  
活かしたい

ファインチューニングは  
無理

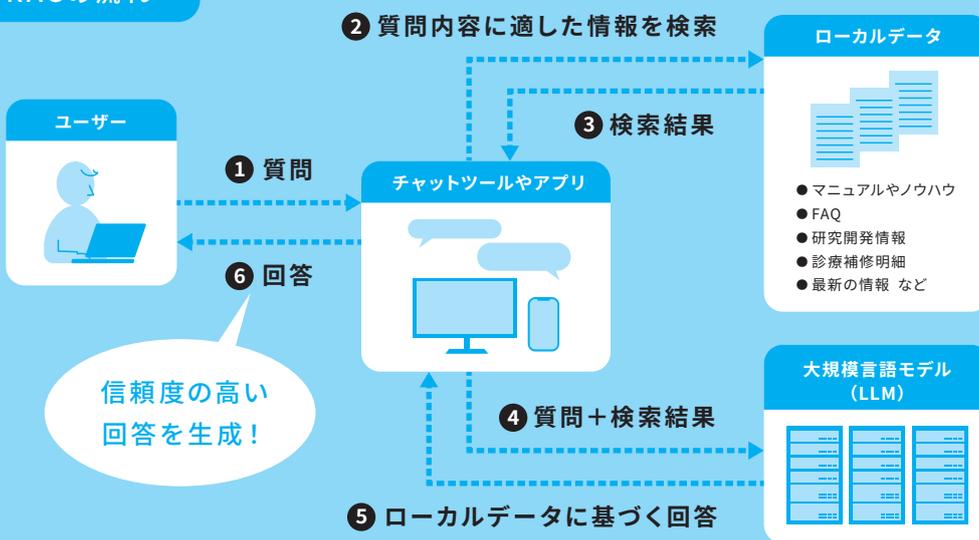
LLMの回答精度が  
いまいち

## RAGの活用で、LLMの回答精度アップ！

### RAGとは

LLM(大規模言語モデル)に、ローカルデータへの検索を組み合わせることで、正確な情報に基づいた回答を生成させる技術です。

### RAGの流れ



### メリット

- LLMが出力する結果を、手軽に最新情報に更新することができる
- 信頼度の高いデータから情報を得るため、誤情報を生成する“ハルシネーション(幻覚)”の抑制につながる

### [ウェビナー] RAGのデモと解説

blueqat社と共催！

オンプレ環境にRAGを搭載する上での「環境構築と仕組みについて」、また「構築後の活用方法やLLMの回答精度の上げ方」、「企業内事例」をblueqat社 代表の湊雄一郎氏にお話しいただきます。ぜひご参加ください。  
開催日時: 2024年10月9日(水) 18:00 ~ 19:00 詳しくは、お問い合わせください。

お問い合わせ

どんなことでもお気軽にご質問・ご相談ください！

弊社サイトのお問い合わせフォームへご入力いただくか、下記までご連絡ください

メール

vt-sales@v-t.co.jp

TEL

03-6823-6789

(受付時間  
平日10:00~17:00)

お問い合わせ  
フォーム▶



www.v-t.co.jp



ビジュアルテクノロジー株式会社

〒111-0052 東京都台東区柳橋2-1-10