

SambaNovaで作る 爆速AIエージェント

～エージェント時代の最強の武器は「速さ」だった～

株式会社スリーアップ・テクノロジー
代表取締役 三上 典秀

いきなりですが、 作ってきました。



皆さん、SambaNovaで
AIエージェントが動いたら
面白いと思いませんか？

⚠ WARNING: HIGH VOLUME / 爆音注意

🎧 SambaNovaの高速推論中だけ、自作の非公式テーマソングが流れます。

- エージェントが超高速で推論・コーディングしている間、ただ画面のロードを見つめるのはもったいない。
- 推論中（AI稼働中）のみ、画面背景で専用のサイバーパンクMVとテーマソングが流れる特別仕様。
- 圧倒的な「推論スピード」と「音楽」のセッションをお楽しみください。

STATUS: LIVE DEMO

PROCESSING...
SPEED: MAXIMUM | RDU: SN50 ACTIVE

スリーアップ・テクノロジー
THREE UP TECHNOLOGY

SambaCloud Agent

SAMBANOVA POWERED | 400+ TOKJS | ZERO LIMIT

Ultra Fast | 400+ tok/s | SambaCloud Agent

NEW TASK | SCHEDULE

- 今日の日付のニュースを5つインターネットで...
2025/04/05 22:10
- イランは今もインターネットが繋がっていな...
2025/04/05 22:07
- イラン攻撃にClaudeが使われたらしいが、W...
2025/04/05 22:01
- 2026年に日本の企業がサイバー攻撃を受けた...
2025/04/05 21:49
- この1週間のアンソロピックのクロードのニュ...
2025/04/05 21:42
- トランプ大統領の直近1ヶ月の発言を時系列で...
2025/04/05 21:32
- ホルムズ海峡の直近1週間の状況を、ニュース...
2025/04/05 21:11

/Users/mikami/Downloads 出カトークン (今回) 179 / 10,000 コンテキスト使用量 14,575 / 32,768 100%

三上 16:39:11
このコンピューターのスペックと、メモリの使用状況とかSSDの使用状況を調べて、サイバーパンク風にHTMLでまとめてグラフ表示して、できたらブラウザ表示。

ツール実行 16:40:24

```
シェル実行:  
echo "=== CPU ===" && lscpu | grep -E "Model name|CPU|Thread|Core" && echo -e "\n=== Memory ===" && free -h && echo -e "\n=== Disk ===" && df -h | grep -E "" | dev | Filesystem
```

```
シェル実行:  
systeminfo | findstr /C:"Processor" /C:"Total Physical Memory" /C:"Available Physical Memory" /C:"Disk"
```

```
シェル実行:  
cmd.exe /c "systeminfo"
```

```
Python実行:  
import platform ...
```

```
シェル実行:  
pip install psutil
```

```
Python実行:  
import platform ...
```

書込:
system_status_cyberpunk.html

ブラウザで表示:
system_status_cyberpunk.html

【Task 1】：挨拶代わりに – フォルダー内のファイルをジャンルごとに自動整理。

【Task 2】：本命の研究 & コーディング – 2026年の日本企業のサイバー攻撃被害状況とリスクを調査。

Flow: Google API検索 → 思考プロセス → サイバーパンク風HTMLレポート自動生成 → ブラウザ表示

限界突破 / SambaNova (日本語版)



SambaNova - Limit Break (English Ver.)



SunoとGeminiを使用して作成した自作の非公式MVです。YouTubeで限定公開中。(本資料はconnpassにアップロードします)

なぜ「製造業」の私が、 こんな爆速エージェントを 作ったのか？

現場力
×
革新技術

三上 典秀

事業内容：製造業向けの産業用IoT・AI装置の開発・設置

現場実績：ファクトリーオートメーション（FA）領域で
26年以上の経験。50設備以上のAI/IoT化実績。

技術実績：2025年全日本AIハッカソン 優勝 🏆



製造業のリアル

⚠ 「情報漏洩が怖い!クラウド (API) 利用は原則禁止!

厳格なクラウド制限がイノベーションを停滞させている



危機感: クラウドを使えない製造業は、
この劇的な進化から完全に取り残される。

世の中の進化

» AIエージェント時代 (Claude Code等) への突入。

業務ソフトウェアが「誰でも簡単に内製化」できる時代の到来。



2026年の希望:

Qwen3.5やGemma4など、エージェントとして実用レベルのオープンソース・ローカルLLMがついに登場。

ローカルLLMの現実:

しかし、ハイエンドGPUを使用しても「推論の待ち時間」が長すぎる。



• エージェントはタスク達成のために、このループを何度も自動で回す。

• 1回の推論の遅延が、ループ全体では致命的な待機時間 (タイムロス) となる。



結論: 「AIの遅さ」が、システム全体のボトルネックになる。

解決策：SambaNovaがもたらすパラダイムシフト

SambaNova オンプレミス環境 (SN50 RDU) =
クラウド水準のAIをセキュアな自社環境で実現。

gpt-oss-120b: > 600 tokens/sec

DeepSeek-V3.1: up to 200 tokens/sec

IMPACT

クローズドなクラウドサービスと比較しても「体感10倍速」の圧倒的推論速度。待ち時間が完全に消滅。エージェントの思考ループが一瞬で完結する。



「AIが速すぎて、人間がボトルネックになる世界が来る」



日本の製造業を、世界に勝てる製造業へ。

- 2026年、サイバー攻撃は間違いなく激化する。AIを使った脅威には、AIで対抗するしかない。
- だからこそ、セキュアなオンプレミス環境で超高速推論を実現する「AIサーバー」が、これからの製造業の絶対的な必需品となる。

**「私は製造業に
SambaNovaの高速推論を導入し、
必ず成功事例を作ります！」**

本日はありがとうございました。