

ITトレンド調査

～生成AIソリューションにおけるプラットフォーム動向～

2024年10月

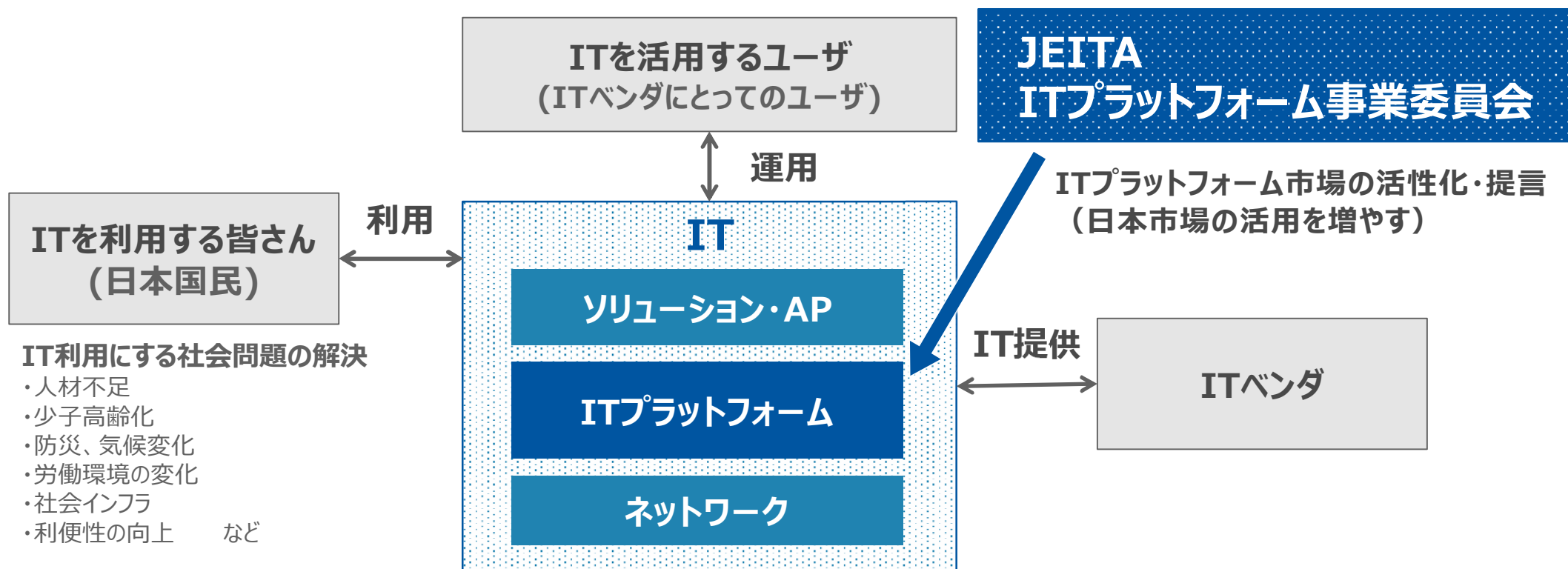
ITプラットフォーム事業委員会
プラットフォーム市場専門委員会
プラットフォーム企画専門委員会
プラットフォームグリーンIT専門委員会
株式会社富士キメラ総研

委員長 竹田 薫
委員長 高橋 謙之
委員長 飯田 靖
委員長 久米 真
辻田 洋佑

JEITA 一般社団法人
電子情報技術産業協会

ITプラットフォームと当事業委員会の関わり

ITプラットフォームに関する製品動向・市場動向などを把握し、当該事業の普及啓発の推進に努める。また、共通課題などに関する必要な施策を検討し、関係先に提言・要望するなどして、業界の活性化を図る



ITプラットフォーム事業委員会の活動

ITプラットフォーム事業委員会

プラットフォーム市場専門委員会

プラットフォーム企画専門委員会

プラットフォームグリーンIT専門委員会

<基本方針>

成長するAI/IoT、クラウドに重点を置き、グローバルとの比較にも取り組む。
継続する活動においても、内容を見直して積極的に変革して行く

攻めのIT、ITプラットフォームの社会への貢献度、出荷統計の調査・公表、
電子情報産業の世界生産見通し調査への協力ほか

ITプラットフォームやクラウド・エッジの将来の姿、DX化におけるITプラット
フォーム使用状況調査、特定分野のユーザ訪問調査、次世代IT技術分
野の事例研究ほか

ITプラットフォームに対する省エネの課題抽出と提案、データセンター等の
現地視察調査、次期省エネ法などのための関係省庁への協力ほか

■ 共通事項

IT関連記者への発表、CEATECコンファレンスでの講演、各種報告書やホームページでの成果公表ほか

ITプラットフォーム事業委員会の構成

JEITA

【委員会参画会社(6社)】(五十音順)

インテル(株)

日本アイ・ビー・エム(株)

(株)日立製作所

沖電気工業(株)

日本電気(株)

富士通(株)

- インテル(株)は、サーバ出荷統計には参加していない
- レノボ・ジャパン(同)は、委員会には参加していないが、サーバ出荷統計は参加

本日の発表 <ポイント>

ITトレンド調査

～生成AIソリューションにおけるプラットフォーム動向～

I. プラットフォーム市場動向

〔プラットフォーム市場専門委員会〕

- 2023年度のIAサーバ出荷実績の報告

II. 国内のサーバ消費電力量の動向

〔プラットフォームグリーンIT専門委員会〕

- サーバの消費電力量の変化と仮想化利用による省エネ化効果

III. 生成AIソリューションにおけるプラットフォーム動向調査

〔プラットフォーム企画専門委員会
富士キメラ総研〕

- ITユーザトレンド調査(2023年度発表)より、生成AIの利用動向
- 生成AIユースケース、生成AIプラットフォーム環境の現状および今後の展望
- データセンター事業者における生成AIソリューション対応に向けた取り組みなどの調査

<プラットフォーム市場専門委員会 発表>

I. プラットフォームの市場動向

～ 出荷統計から見てくるITプラットフォーム市場の現状と展望 ～

サーバ出荷統計の特色

- JEITAのサーバ出荷統計は、下記のような特色があります。
 - 参画会社各社の実績データそのものを集計したもので予測・推測は一切含まれていない。
 - 過去から継続してデータを収集し、統計を実施。
 - 月単位でのOS機種別、価格クラス別での自主統計を実施し、四半期単位で機種別の実績を公表。
 - 半期単位でOS機種別、価格クラス別、産業別の出荷実績を算出し、公表。
 - 出荷金額はハードウェアシステム（本体、磁気ディスク装置等の周辺機器）および基本ソフトウェア関連の範囲とし、ユーザ固有のアプリケーションソフトウェア金額や保守料金は含まない。

2023年度 I Aサーバ出荷実績

- I Aサーバにおいて、出荷台数は減少したが、金額は前年度を大幅に上回った。
- 価格帯別にみると、「300万円以上」が出荷台数および金額ともに前年度を大幅に超過した。
- 一方、出荷台数においては、「25万円～50万円未満」が前年度比88%、「25万円未満」が69%と低い水準であった。

出荷台数 (台)

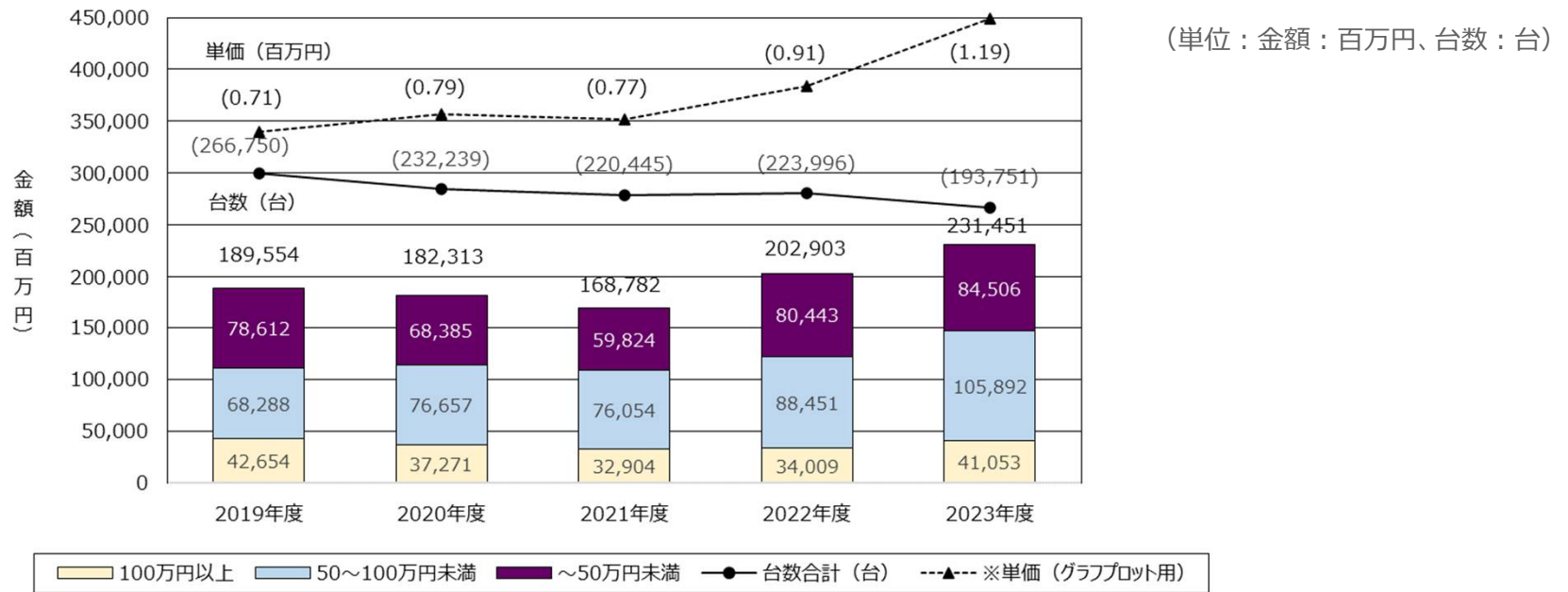
		年度合計						
		上半期			下半期			
		第1四半期	第2四半期	第3四半期	第4四半期	第3四半期	第4四半期	
出荷台数 (台)		48,629	53,114	101,743	45,736	46,272	92,008	193,751
下段：前年比 (%)		98%	92%	95%	75%	82%	79%	86%
価格帯別	300万円以上	198	677	875	98	122	220	1,095
		190%	410%	325%	88%	103%	96%	219%
	100万円～	1,837	2,128	3,965	2,685	3,453	6,138	10,103
	300万円未満	43%	81%	57%	112%	138%	125%	85%
	50万円～	12,822	17,638	30,460	13,926	14,821	28,747	59,207
	100万円未満	115%	116%	116%	82%	103%	91%	102%
	25万円～	20,004	21,280	41,284	19,314	18,707	38,021	79,305
	50万円未満	116%	89%	100%	75%	80%	77%	88%
25万円未満	13,768	11,391	25,159	9,713	9,169	18,882	44,041	
	83%	73%	78%	64%	57%	60%	69%	

出荷金額 (百万円)

		年度合計						
		上半期			下半期			
		第1四半期	第2四半期	第3四半期	第4四半期	第3四半期	第4四半期	
出荷金額 (百万円)		54,968	61,698	116,666	58,572	56,213	114,785	231,451
下段：前年比 (%)		137%	121%	128%	103%	102%	103%	114%
価格帯別	300万円以上	4,378	3,053	7,431	897	1,183	2,080	9,511
		348%	245%	297%	79%	69%	73%	178%
	100万円～	6,132	7,014	13,146	8,758	9,638	18,396	31,542
	300万円未満	80%	104%	91%	120%	138%	129%	110%
	50万円～	21,530	29,792	51,322	27,023	27,547	54,570	105,892
	100万円未満	140%	138%	139%	104%	108%	106%	120%
	25万円～	10,448	12,227	22,675	12,262	10,251	22,513	45,188
	50万円未満	141%	113%	124%	94%	98%	95%	108%
25万円未満	12,480	9,612	22,092	9,632	7,594	17,226	39,318	
	147%	91%	116%	105%	73%	88%	102%	

I Aサーバの平均単価推移

- 平均単価は 1.19百万円となり、年々上昇傾向であったが、今回初めて100万円を超える結果となった。
- 単価上昇の要因として、データセンターでのI Aサーバの利用が高まり、HCI（HyperConverged Infrastructure）利用や、AI や科学技術用途向けのGPGPU（General-Purpose computing on Graphics Processing Units）搭載、その他オプシオン機器搭載により、サーバ構成／スペックの拡張によるものと推察する。



2024年度以降の見通しについて

- 生産性向上や価値創造などへの対応から新たに生まれるワークロードに最適に対応するべくサーバ需要も堅調に推移することが想定される。
- さらに、生成AIの急速な普及により、業務適用への流れも加速すること、また、クラウドを活用したシステム・サービスの拡大に対応したデータセンタ構築・増強の動きが活発に推移することが想定される。
- これらの動きに対応した下記の分野・領域において、ITプラットフォームのさらなる需要の拡大が期待できる。

- **ビッグデータの高速解析や人工知能（AI）による新たな価値創造への取り組み**
- **サーバの高速処理による、よりリアルタイムな在庫管理や工場ラインの工程管理や検査管理**
- **5Gなど通信インフラの整備やIoTデバイスの浸透に伴うデータ量の増加など市場変化への対応**
- **クラウドを活用したシステム・サービスの拡大に対応するデータセンター構築・増強**
- **社会や市場からの要請による、高度なサイバーセキュリティへの対応**
- **システム運用効率化に向けたサーバ統合・仮想化からシステム統合への取り組み拡大**

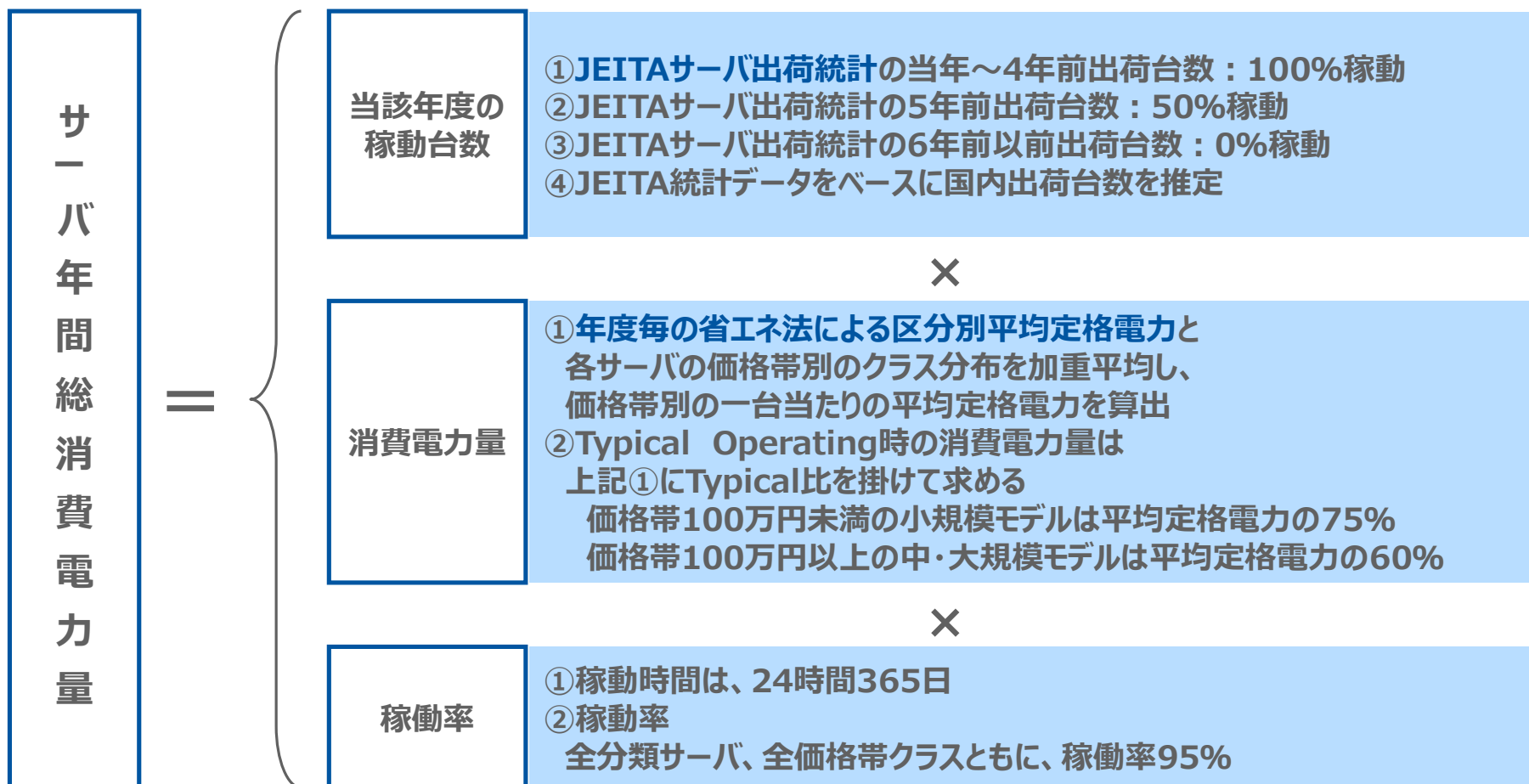
<プラットフォームグリーンIT専門委員会 発表>

Ⅱ. 国内のサーバ消費電力量の動向

サーバ消費電力量動向の特色

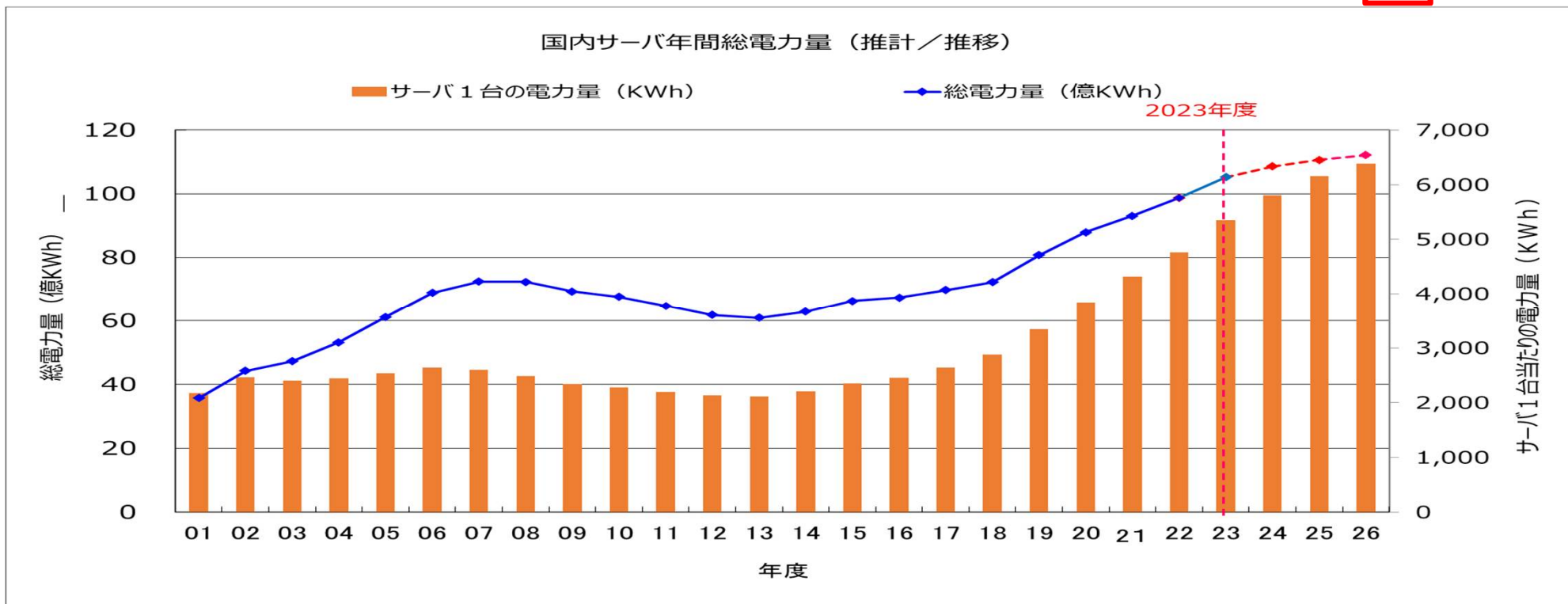
- JEITAのサーバ消費電力量動向は、下記のような特色があります。
- 参画会社各社の出荷実績データを元に、推移状況を把握して、国内総量を推計。
- 過去から継続してデータを収集し、統計を実施。
- 物理サーバの年間総消費電力量推移と、物理サーバ上における仮想システムの稼働動向、仮想化による省エネ効果の推移を算出。

サーバ年間消費電力量の試算基準



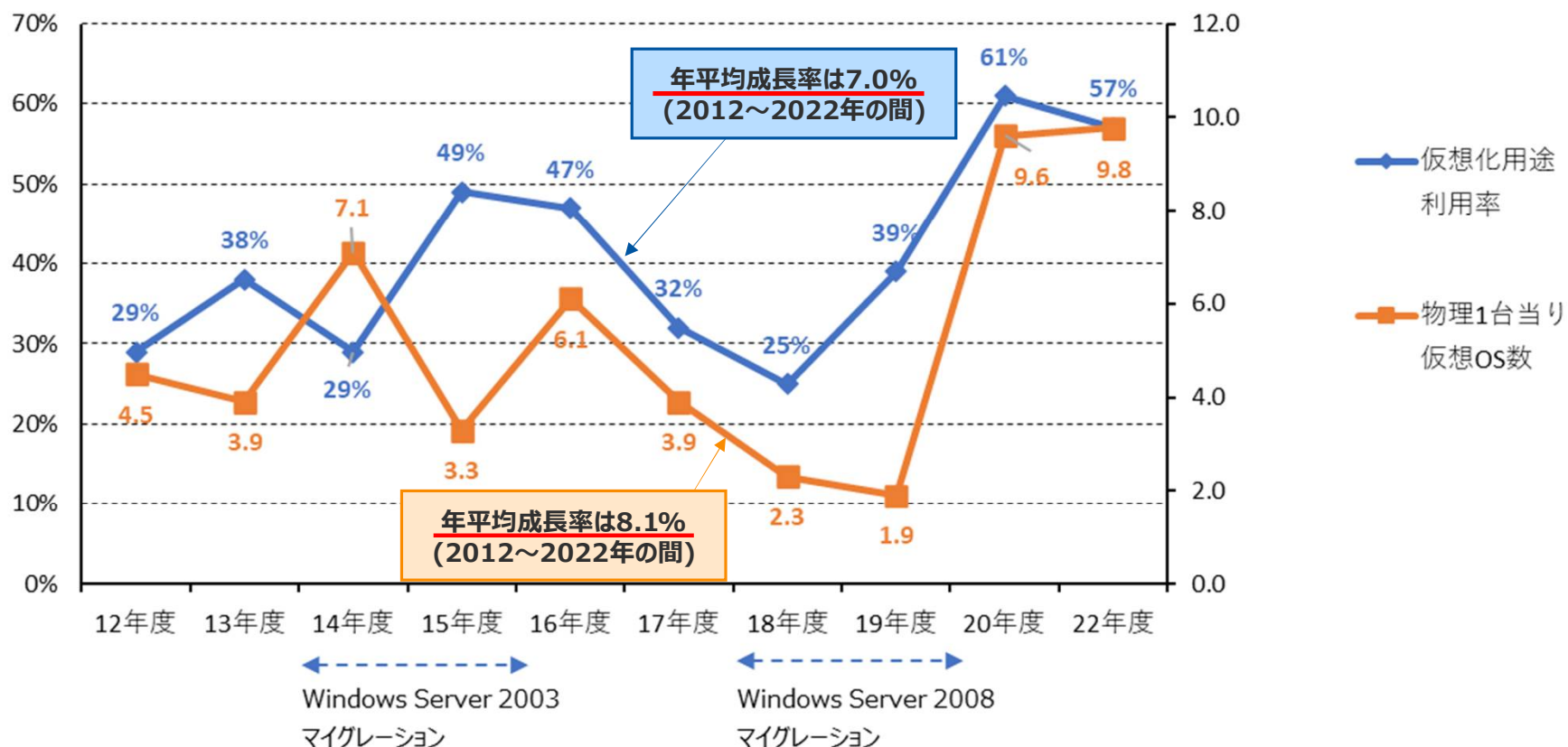
サーバ年間消費電力量の推移

年度	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
物理台数	総電力量 (億kWh)	36	44	47	53	61	69	72	72	69	68	65	62	61	63	66	67	70	72	81	88	93	99	105	109	111	112
	稼働台数 (万台)	164	180	197	217	241	261	278	291	296	297	295	289	288	285	283	275	264	251	241	229	216	208	197	187	180	176
	サーバ1台の電力量 (kWh)	2,178	2,462	2,403	2,446	2,539	2,638	2,603	2,489	2,338	2,277	2,194	2,134	2,112	2,208	2,346	2,450	2,639	2,876	3,343	3,843	4,312	4,760	5,347	5,802	6,161	6,379



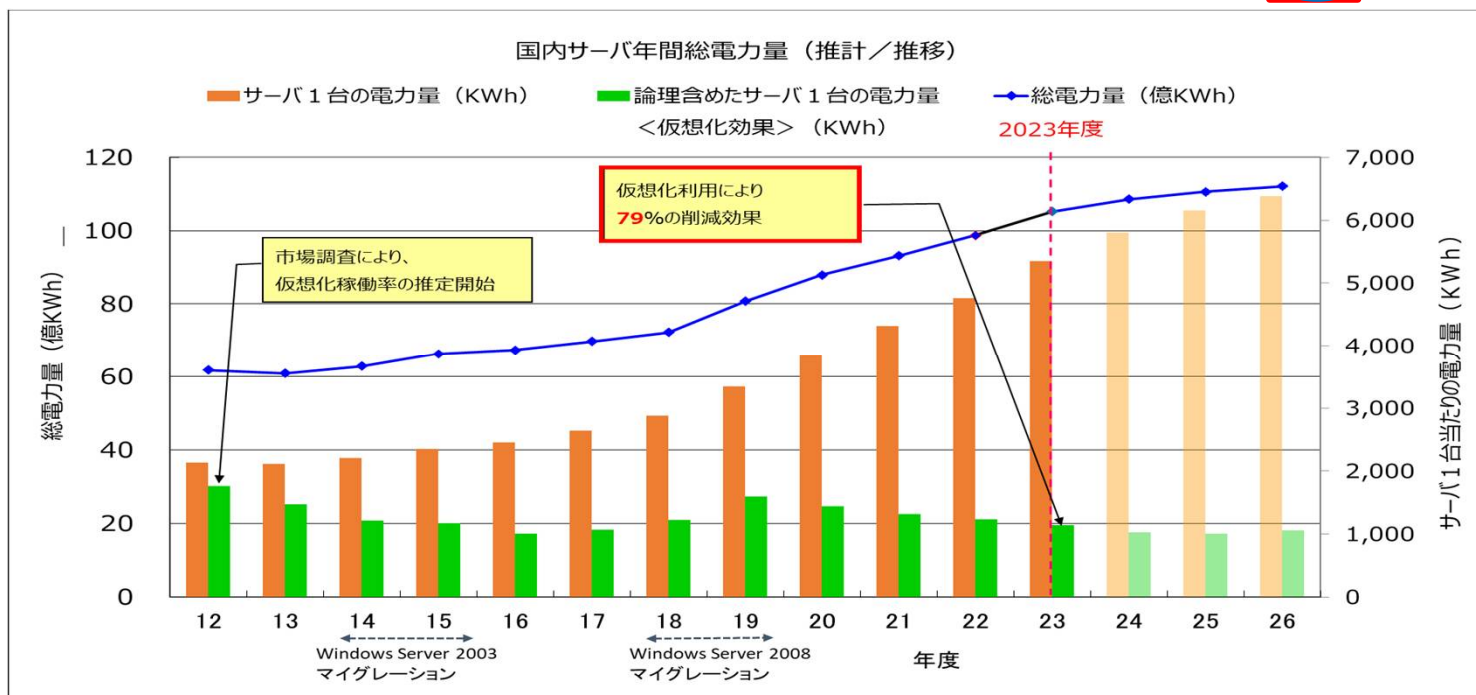
物理サーバ上における仮想システムの稼働動向

(2023年実施のITトレンド調査より)



サーバ年間消費電力量の推移（仮想化による省エネ効果）

	年度	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
物理台数	総電力量（億KWh）	62	61	63	66	67	70	72	81	88	93	99	105	109	111	112
	稼働台数（万台）	289	288	285	283	275	264	251	241	229	216	208	197	187	180	176
	サーバ1台の電力量（KWh）	2,134	2,112	2,208	2,346	2,450	2,639	2,876	3,343	3,843	4,312	4,760	5,347	5,802	6,161	6,379
仮想化効果	論理含めた稼働台数（万台）	351	414	519	567	668	655	591	508	612	707	802	920	1,053	1,098	1,061
	論理含めたサーバ1台の電力量 <仮想化効果>（KWh）	1,759	1,471	1,209	1,169	1,009	1,065	1,222	1,590	1,436	1,318	1,231	1,143	1,031	1,008	1,057
	電力削減効果（%）	18%	30%	45%	50%	59%	60%	58%	52%	63%	69%	74%	79%	82%	84%	83%



2024年度以降の見通しについて

- CPUのコア数増加とマルチプロセッサ化が進み、1台当たりの物理サーバ消費電力量は今後も増加していく傾向。
- サーバの仮想化用途利用、および物理サーバ1台当たりの仮想OS数増加により、79%の電力量削減効果が得られていると推定。
- 1台当たりの物理サーバ消費電力量の増加、およびITプラットフォームの導入拡大の傾向から、サーバ全体の年間消費電力量は今後も増加すると想定される。
- サーバ消費電力量の削減には、高効率な冷却技術での冷却電力削減など、サーバラックあたりの省エネ効果を増やす取り組みが必要。
- 今年度昨今の生成AI需要の高まりから本用途に利用されるGPUサーバに着目しプラットフォーム企画専門委員会と協調し調査を実施した。GPUサーバは1台当たりの消費電力量が多いが、当面データセンター総消費電力量に占める割合は低く推移する見込みである。しかしながら、今後汎用サーバを上回る伸び率で導入が進むと予想されることから注視は必要と考える。

<プラットフォーム企画専門委員会・株式会社富士キメラ総研 発表>

Ⅲ. 生成AIソリューションにおけるプラットフォーム 動向調査

(利用か所有か。生成AI(LLM)システムに関わる勘所)

■ ITユーザトレンド調査（隔年実施）

- ・一般企業および公共・公益法人7,000余りの団体にインターネットによるアンケート
2023年調査では有効回答816件(有効回答率10.5%)
- ・情報システム部門だけでなく、役員クラス(CIO等)、業務部門(管理・事業)も含む調査
過去年の定点観測の項目は、ITシステム部門の回答に絞って分析

2023/8
ITユーザトレンド調査
【含むAI】

2024/1
注目分野に関する
動向調査報告
【生成AI】

■ 生成AIとITプラットフォームに関する調査（2024年7月実施）

- ・2023～2030で関わる市場は2桁成長。その旺盛な市場におけるITプラットフォームの動向、課題について調査（1月調査報告からの深堀、ユースケースの調査）
- ・クラウドサービスの利用かオンプレミスでの所有か、LLMなどシステムの持ち方や関わるデータセンタでの課題などについて考察

2024/7
生成AIとITプラット
フォームに関する調査

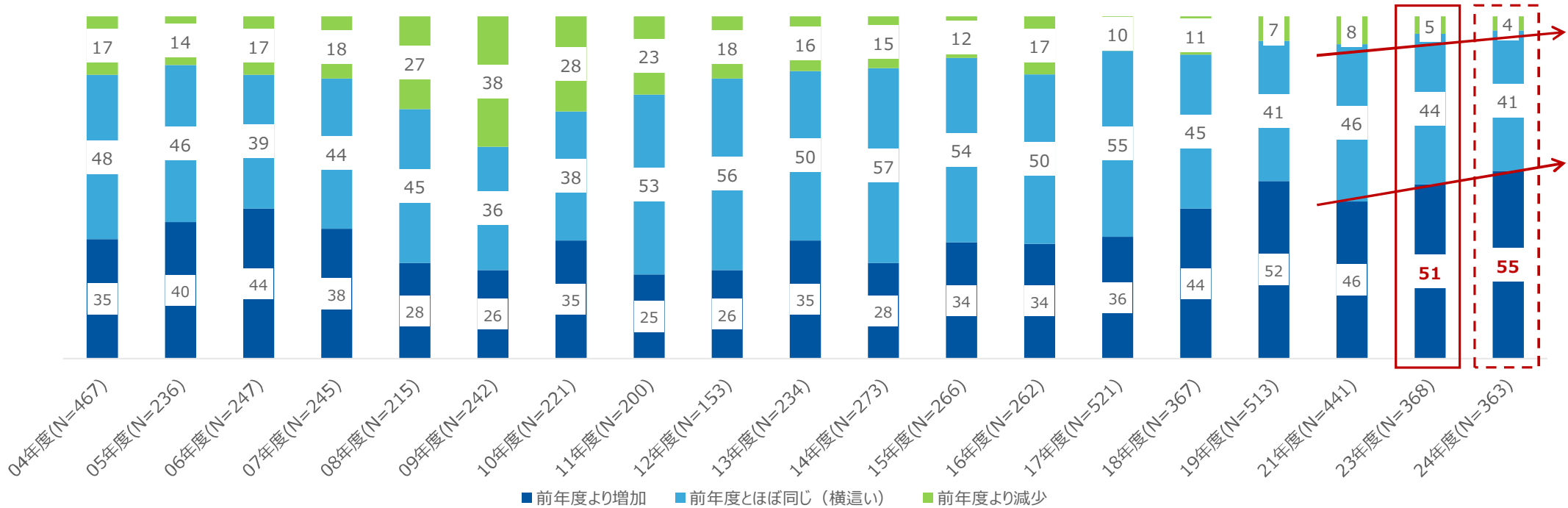
LLM : Large Language Models (大規模言語モデル)

IT投資の動向 ～2023年の推移と2024年の予測

※ 情報システム部門限定

- 2023年度は、「前年度より増加」が過半数、「前年度とほぼ同じ」を含めると過去最高の95%
- 2024年度も、「前年度より増加」が過半数、**IT投資に対する高い意欲**が見受けられる

企業/組織ユーザーのIT投資の推移（情報システム部門回答による） ※24年度は見通し

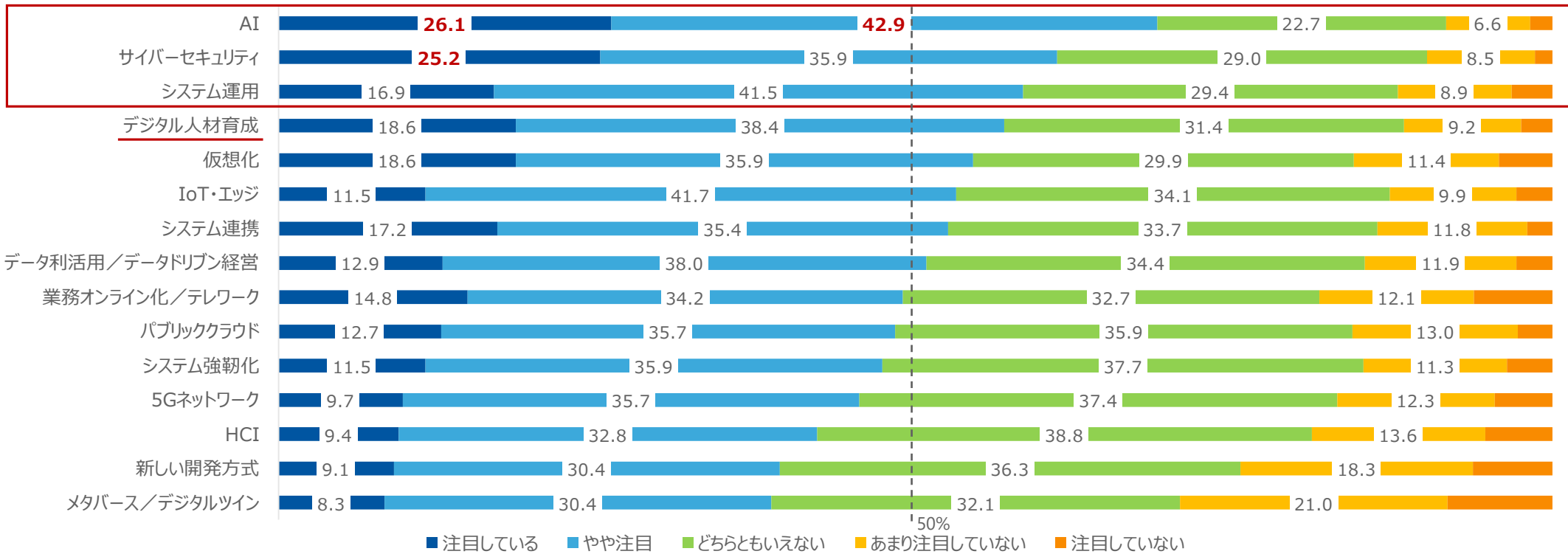


(注) 2019年度調査以降、対象を非情報システム部門にまで拡大したが、上図では「情報システム部門のみ」集計している。

15のテーマ別でみた注目度（全体値）

- 「AI」、「セキュリティ」、「システム運用」の順。8テーマで「やや注目」以上の割合が過半数
- 「AI」に対する注目度が高まっている（「やや注目」以上が7割）

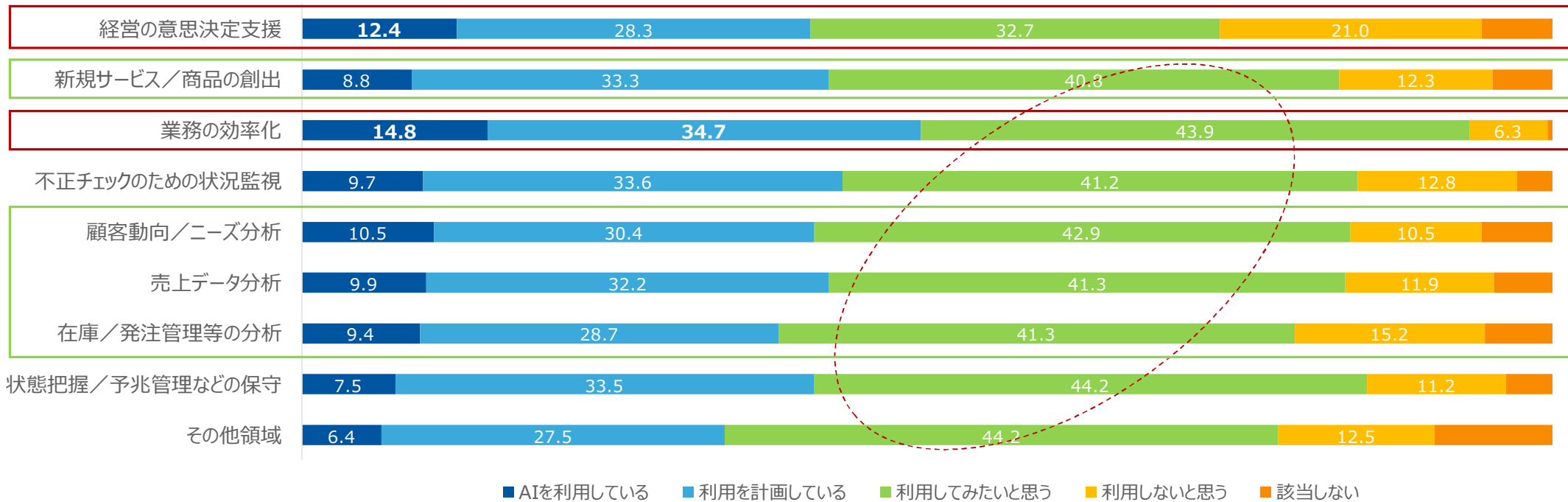
「やや高い」以上が多い順（全体値）（N=816）、単位：%



用途別でみたAIの利活用意向（利用目的）

- AIの利活用意向で「利用している」が多く見られるのは、「**業務の効率化**」と「**経営の意思決定支援**」
- 「**新規サービス/商品の創出**」「**顧客動向/ニーズ/売上データ/在庫などの分析**」といった**ビジネス強化に繋げる期待も**
- 各用途とも「**利用してみたいと思う**」が4割を占める。**今後、AIの利活用に向けた動きが進むと予測**

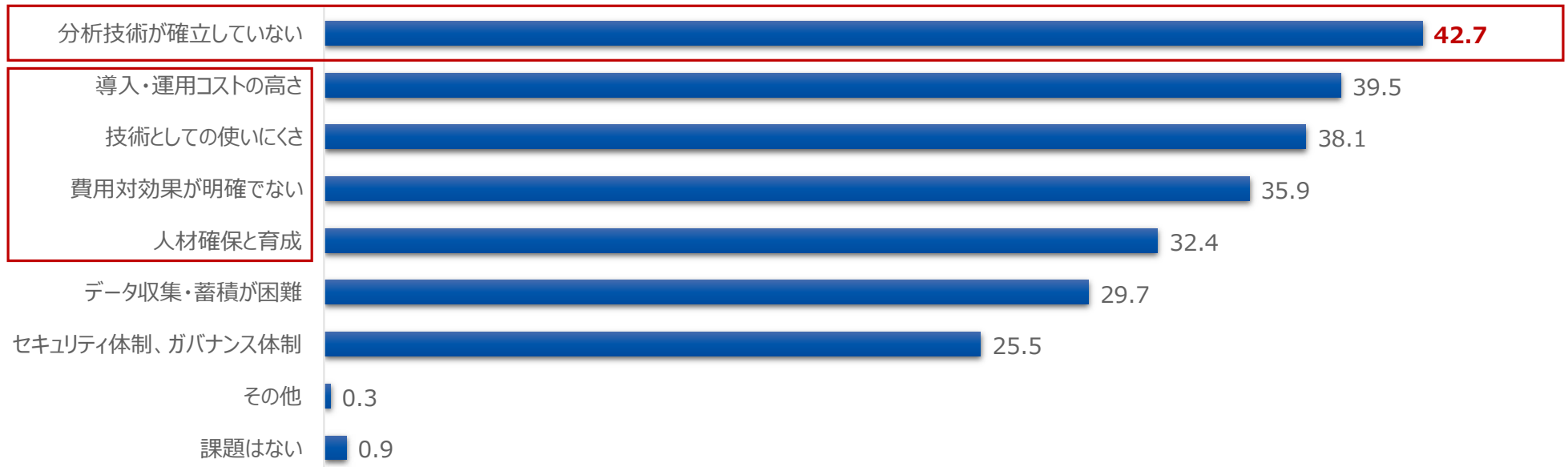
用途別に見たAIの利活用意向（各項目・N=816）



AIを利活用する上での「課題」は何か

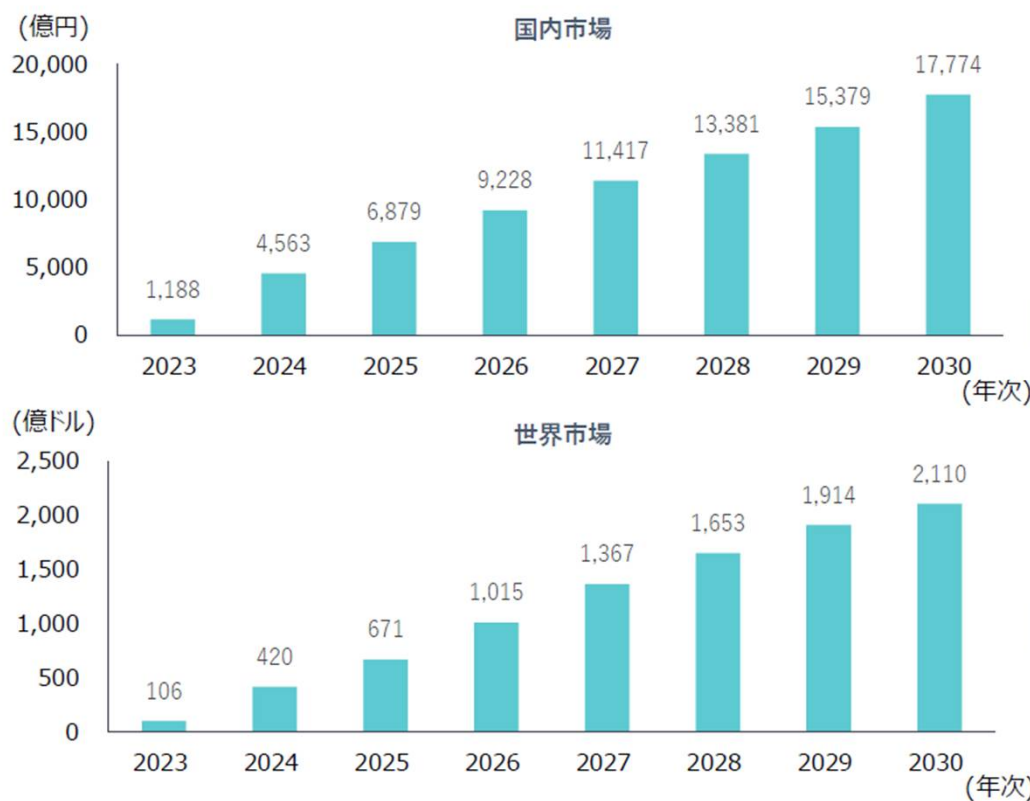
- AIの利活用にあたっての課題感は、「**分析技術が確立していない**」（42.7%）、「**導入・運用コストの高さ**」（39.5%）、「**技術としての使いにくさ**」（38.1%）、「**費用対効果が明確でない**」（35.9%）の順
- 利活用に向けた動きは進んでいるが、**使い易さなどの技術面、コスト、人材確保に課題**

AI技術の利活用にあたっての課題は何か（N=816、複数回答）



国内/世界における生成AI市場規模

■ 生成AIに関わるビジネス市場は2桁成長を望む。特に2024年は前年比約4倍と急成長



生成AI市場は急速に拡大

- ChatGPTの登場により生成AIの業務適用に対する期待値の高まり、個人レベルでの有用性検証が為されていた。
- ベンダー基盤モデルを軸としたプロダクトアウト、システム開発ニーズの高まりから2023年度に市場拡大が本格化。
- 2024年度以降においても、導入企業の拡大および生成AIのさらなる浸透により好調な市場成長が見込まれる。

生成AI活用はグローバルで利活用が進展

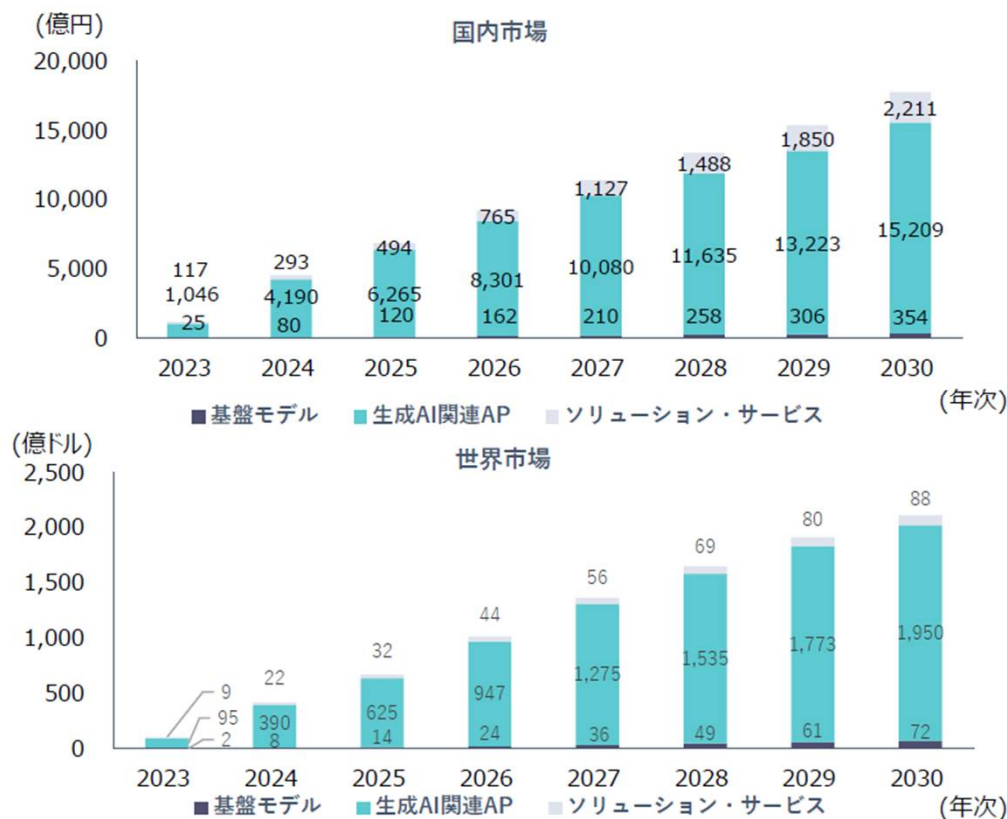
- 世界市場に関しては、国内と比較し基盤モデルを中心とした研究開発が先行していたことから、米国や欧州を中心に市場が拡大。
- アジア圏においても、基盤モデルのユーザーとしてプロダクトへの搭載やモデルを活用したシステム開発ビジネスが進展するものと思われる、高い成長率での推移。

国内においても生成AI市場は伸長

- 導入検討から試験的導入までの期間が短く、2023年度から早期に市場が形成されている。
- 主に、ChatGPTと同様の対話型AIサービスの利用や人材教育やガバナンス整備支援などが先行して導入され、市場が拡大している。

国内/世界における生成AI市場規模推移（ビジネスレイヤー別）

■ 生成AI関連アプリケーションの市場が大きく、対話型AIの活用にとまらず業務特化のニーズも見込む



2023:生成AIニーズの立ち上がり

- 2023年においては、生成AIを活用したプロダクトの急速な普及のほか、生成AI導入にかかわる上流サービスを中心に市場が拡大。
- 周辺市場(生成AI関連アプリケーション、ソリューション・サービス)市場の拡大に伴い基盤モデル市場も伸長。

2025:アプリケーションの急速な普及

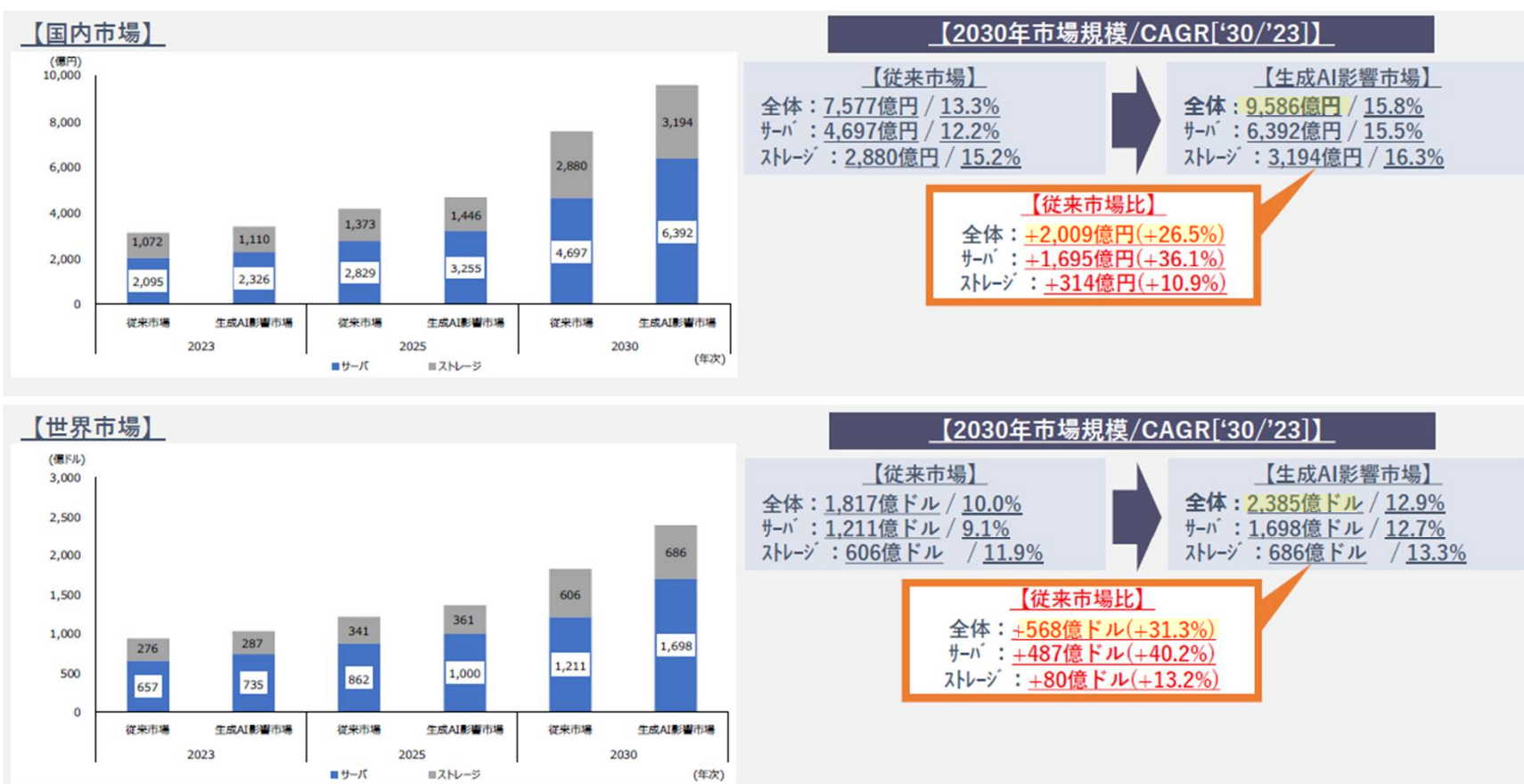
- 2025年度においては、2023年度のChatGPT導入を皮切りに、様々な業務アプリケーションへの搭載のほか、対話型AIを中心に導入が進展。
- 一方、徐々に法人/業務に特化した生成AI活用ニーズが高まるものとみられ、ソリューション・サービスや基盤モデル市場の拡大とともに生成AIの個別開発が増加。

2030:特化型生成AI活用ニーズの拡大

- 2030年度においては、国内/世界問わず汎用的な生成AI活用から、企業または特定業務に特化したAIシステムニーズが高まるものとみられる。
- 一方、ITシステムの内製化が進展している世界市場においては、アプリケーションや基盤モデルを活用した自社適応を行う中、国内ではSIベンダーへの委託による対応が進展するものとみられる。

生成AIが与えるインフラへの影響市場

- ITプラットフォーム市場においても生成AIが市場を持ち上げ +20~30% (GPUサーバ、高性能ストレージ 等)



生成AIに関わるプラットフォームの動向と課題

(2024年7月 調査より)

JEITA

- 具体的なプラットフォームの需要はクラウドから
 - ・調査20社全てが、**現状、パブリッククラウド環境**で利用（主にはGPT基盤とAzure OpenAI活用）
- 今後はRAG、ファインチューニングなどから、**オンプレミスやハイブリッドクラウドでのLLM構築も**
 - ・活用の幅が広がり業務システムとの連携も
 - ・キーとなるのは**取り扱うデータの機密度とセキュリティ**
- LLMシステム構築の課題も
 - ・**GPUサーバ導入・利用技術**（高価、ネットワーク含む構築技術、CUDAなど利用技術）
 - ・**設備負担**（高電力量、発熱処理、高密度実装における耐荷重 等）

利用か？ 所有か？ ⇒ 目的をもって、自社の業務・データの特性、投資対効果から見極めることが肝要

RAG : Retrieval-Augmented Generation (検索拡張生成) CUDA : Compute Unified Device Architecture (NVIDIAが提供するGPU向けの汎用並列コンピューティングプラットフォーム)

<プラットフォーム企画専門委員会・株式会社富士キメラ総研 発表>

Ⅲ. 生成AIソリューションにおけるプラットフォーム 動向調査

(調査結果の詳細)

※本調査はJEITA ITプラットフォーム事業委員会が主体で、富士キメラ総研協力のもと実施いたしました。

生成AIとは

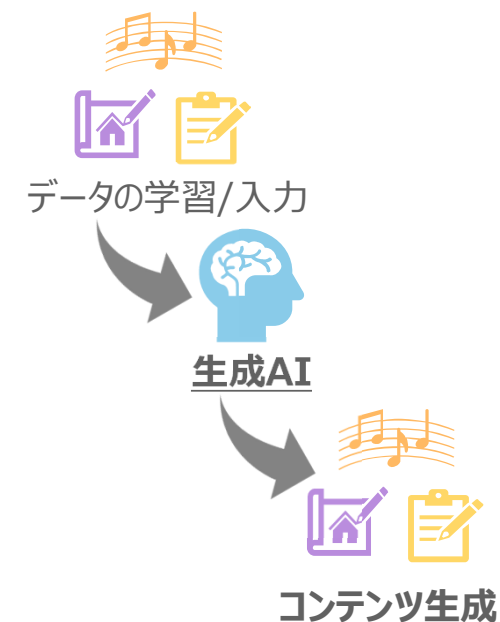
人工知能 : 人間の知的活動を模倣したり、代替したりするコンピュータプログラム

機械学習 : データを学習し、特定のパターンや法則を発見する人工知能モデル

深層学習 : ニューラルネットワークを用いて特徴量の探索を行う機械学習モデル

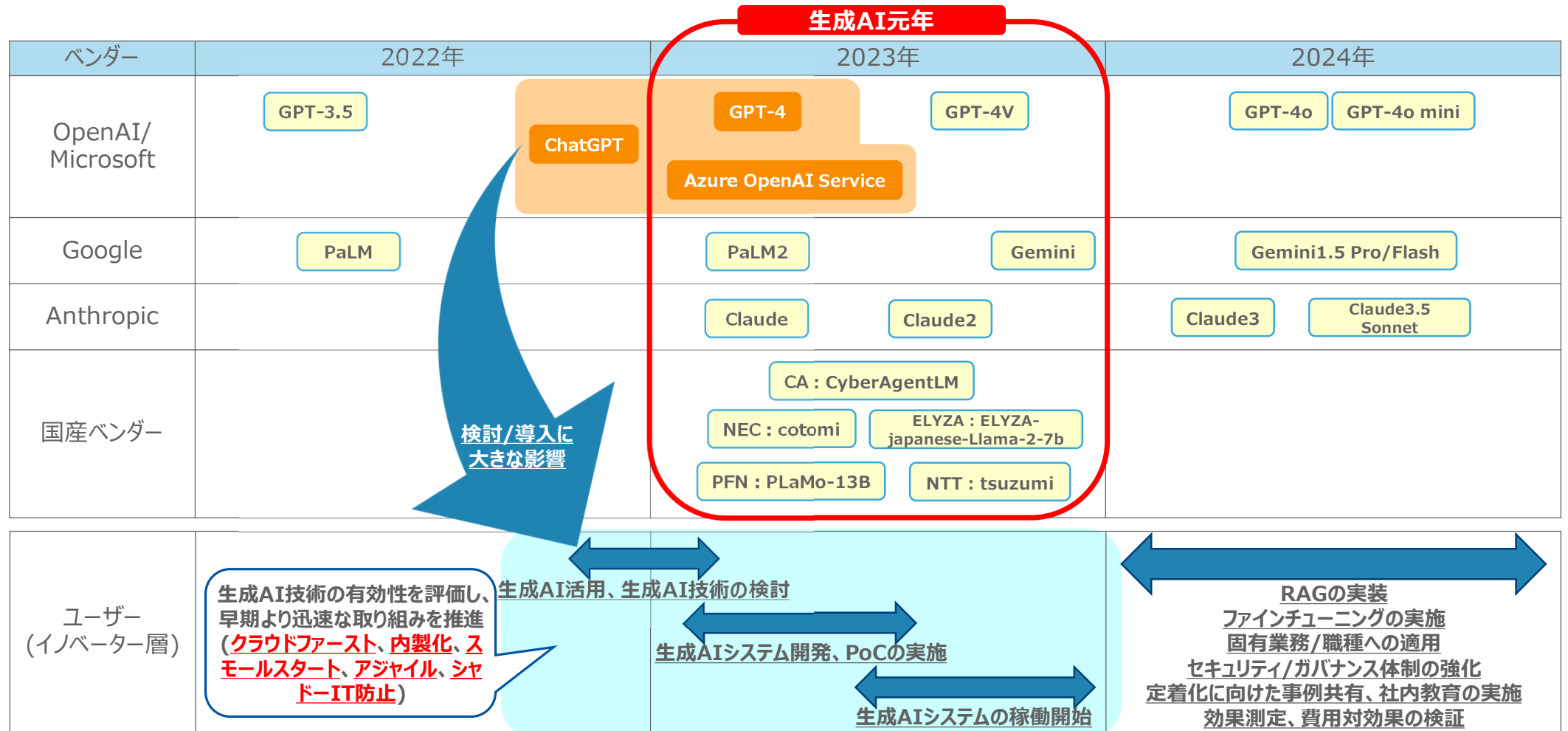
生成AI :

TransformerモデルやDiffusionモデルを用いて、自然言語などの入力に応じた文章/画像/音楽/動画生成を行うことが可能な深層学習モデル

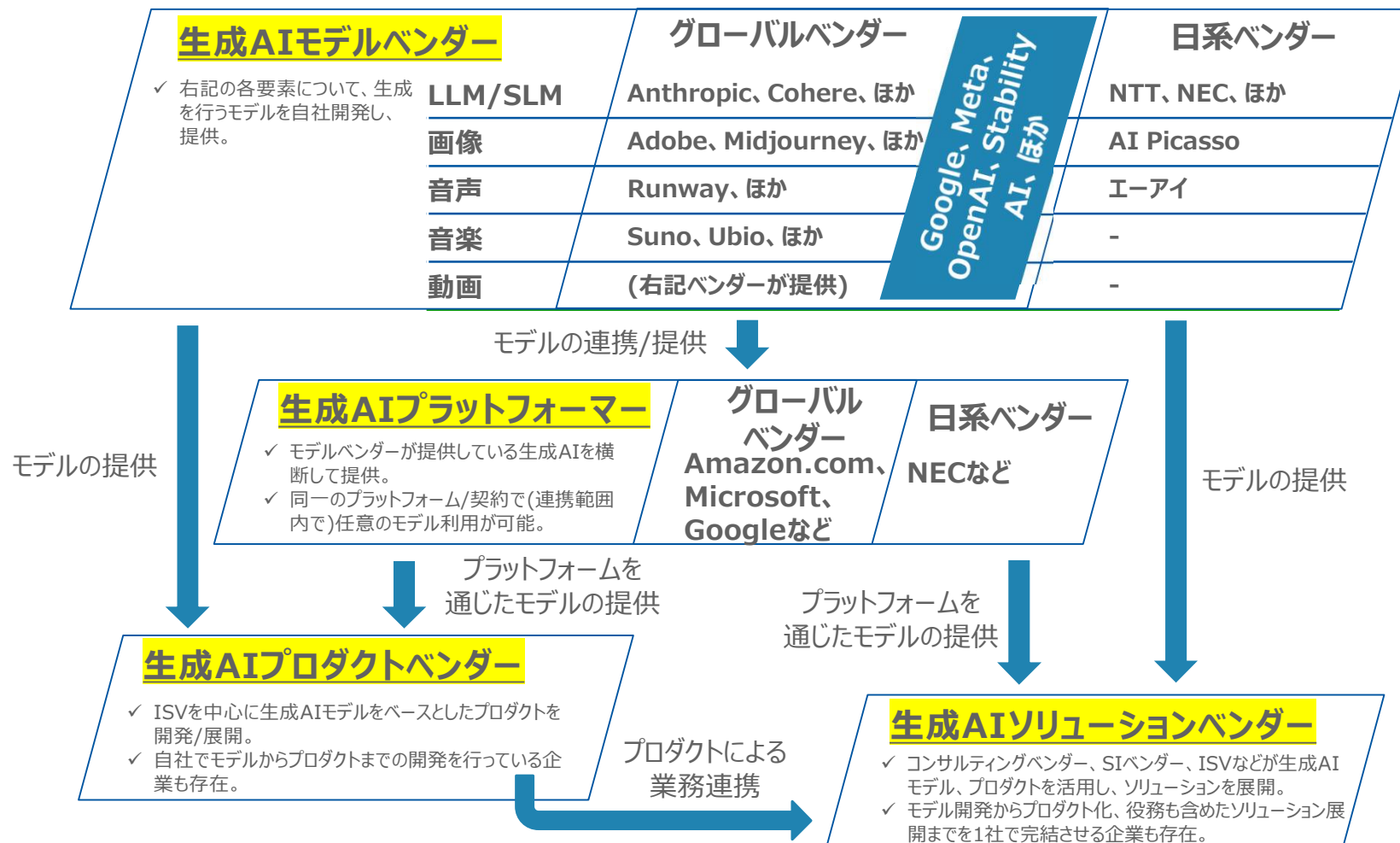


- 深層学習モデルの一種として分類され、文章や画像、動画といったデータをTransformerモデル、Diffusionモデルといった深層学習モデルで学習し、プロンプトに応じて求められた文章や画像、音楽、動画を出力することが可能なAI技術。
- 従来AI : 構造化データを学習/入力にし、特定の目的に応じた認識/分類/予測を行う。
- 生成AI : 文章/画像/音声データを学習/入力し、文章/画像/音声コンテンツの生成を行う

主要ベンダーおよびユーザーの取り組みの変遷



生成AIソリューション市場業界マップ



先行している生成AI(LLM)ユースケース例

情報探索



外部情報の取得

一般的な質問に対し、**Webサイトから最新情報を取得**して回答に反映。**RAGによる特定情報の取得**も実現。

社内情報の検索

日々の業務で活用する**社内文書/マニュアルなどの検索**を支援。**社内ヘルプデスクの業務負担軽減**が実現。

資料作成/ 翻訳/要約



文章作成および文章の翻訳/要約

メール文章やプレゼン資料の作成、文章の翻訳/要約、議事録作成などを支援。**RAGによる過去の会議や資料を活用した文章作成**も実現。

アイデア出し/壁打ち

目的や要望、条件などに応じた**1人でのブレインストーミングの実施**を支援。プロンプト次第で**抜け漏れのないアイデア創出**が実現。

コード生成/ コード補完



コードの生成および生成コードの自動補完

自然言語での指示により、**プログラミングコードの作成および補完を自動化**。**開発時間の短縮**や**品質改善**、**開発業務の平準化**などが実現。

インサイト/ データ解析



データのグラフ化

データの直接入力や表データのアップロードの後、自然言語での指示により、**グラフ作成を自動化**。データの性質に応じた視覚化により**インサイトの引き出し**が実現。

自由記述アンケートの分析

指定したカテゴリーに応じた**分類**やトピックスの**抽出**、**要約**などにより、自由記述アンケート分析を支援。**分析時間の短縮**に加え、**分析業務の平準化**が実現。

生成AI活用への期待



業種固有の有望ユースケース①

製造業



ナレッジ抽出・技術継承、アフターフォローの自動化

- 熟練の技術者が有している暗黙知やナレッジ、技術の抽出・継承が可能となる。
- 製品販売後の不具合対応やFAQ対応などのアフターフォローに関して、一次対応の自動化が可能となる。

金融業



審査業務の自動化、問い合わせ対応の自動化

- 保険加入や銀行手続きで申請者の信用力をスコアリングし、多角的な評価をベースとした審査が可能となる。
- 金融/保険商品や手続きなどに関する顧客からの問い合わせに対し、簡易対応や一次対応の自動化が可能となる。

流通業



最適な在庫管理、広告コンテンツの生成

- 過去の販売データや天候データなど、各種データを参照したうえでの最適な受発注管理、在庫管理が可能となる。
- オンライン/オフライン問わず、広告コンテンツを生成することで、マーケティング業務に資する業務時間の削減が可能となる。

情報通信業



コード生成・補完、運用・保守の自動化

- 設計書や自然言語などによってコードを生成・補完することで、生産性の向上が可能となる。
- 障害や不具合発生時の問題の特定から復旧に至るまでを自動化することで、AIOpsの実現が可能となる。

業種固有の有望ユースケース②

社会インフラ業

建築物のデザイン生成、施工管理の効率化

- 建築物のデザインや設計書の生成により、設計・開発業務に関わる時間短縮が可能となる。
- 施工管理に関わる紙ベースでの入出力業務や関係各所のスケジュール調整などを生成AIによって効率化することで、業務負担低減が可能となる。

ヘルスケア業

問診内容の要約・診療案作成、創薬・研究開発

- 問診票の内容の要約、過去の診療データをベースとした診療案を作成することで、診療業務の効率化が可能となる。
- 過去の公開論文や長年蓄積された研究データをベースとした基礎研究や標的分子探索などにより、これまでより短期間での創薬が可能となる。

公共/文教

申請受付業務の自動化、テスト採点の自動化

- 各種申請に伴う相談内容をベースに適切な申請方法の案内や申請サポートなどを自動化することで、申請受付業務の効率化が可能となる。
- 記述式問題や小論文などの採点を生成AIが行うことで、一律な採点基準のもと公平的な採点が可能となる。

生成AIプラットフォーム環境の現状および今後の展望

現状

クラウド環境の選択が主流

【生成AIシステム検討開始当時の選択肢】

- イノベーター層によって生成AIシステムの検討が開始された当時、基盤モデルの選択肢は限られ、提供形態としてもクラウド環境を前提としているものが中心。

【クラウドの特性との親和性】

- イノベーター層が生成AIシステムの稼働開始を目指す中、迅速かつ低コストでスモールスタートでの開始が可能なクラウド環境は親和性が高い。

【ハイパースケラーによるサービス拡充】

- ハイパースケラーの提供するクラウドサービスの利用が一般化する中、ハイパースケラーが生成AIプラットフォームの立場でのビジネスを推進。

今後

適材適所での環境選択が進展

【データプライバシー/セキュリティの強化】

- RAGやファインチューニングなどに取り組むにつれ、取り扱いデータが増加。中には機密データが含まれるケースもあり、セキュリティの観点からオンプレミス環境も選択肢に挙がる。

【生成AIニーズの多様化】

- 生成AIの普及に伴い、ユーザー側の生成AIニーズも多様化(基盤モデルのカスタマイズ、基盤モデルの使い分け、低遅延性、など)し、生成プラットフォーム環境の適材適所ニーズが増大。ベンダー側のサービス/提供形態の拡充も進展(基盤モデルをインストールしたワークステーションの導入など)。

マジョリティ層における生成AIの普及

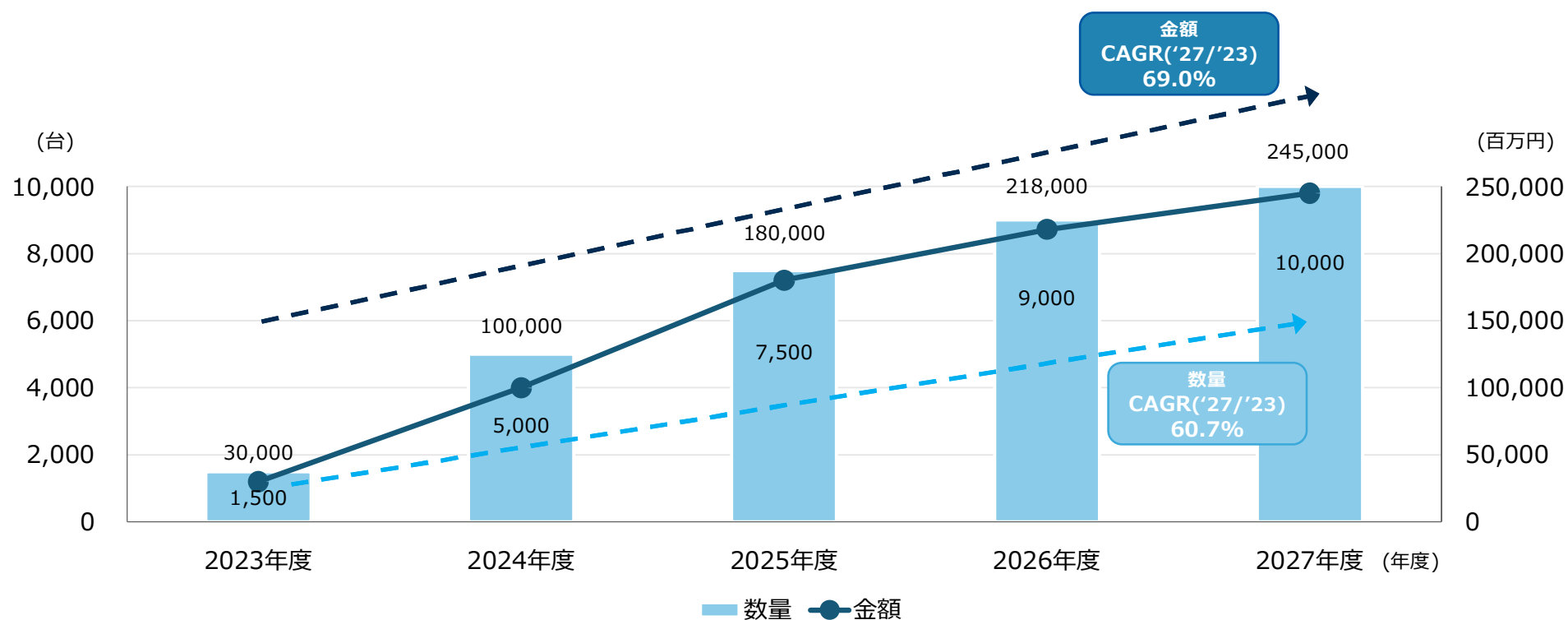
RAGやファインチューニングの進展

基盤モデルの選択肢拡充

基盤モデルのマルチモーダル化

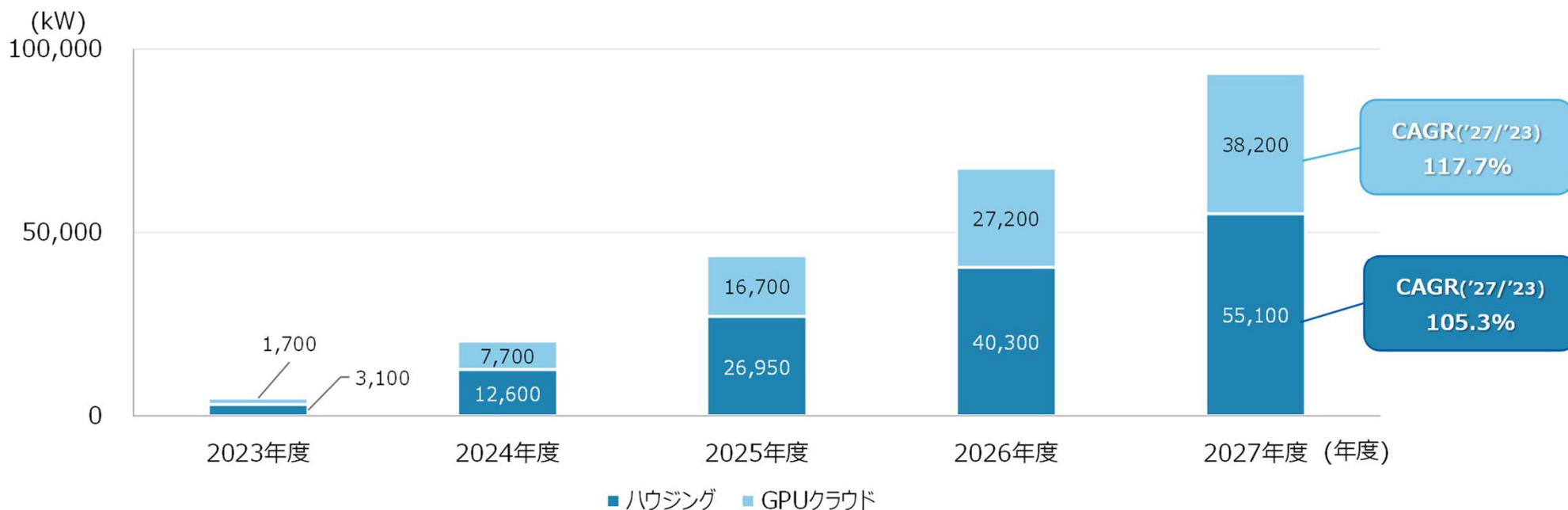
GPUサーバー国内市場規模推移

- 大学や公共の教育/研究機関、製造業の研究機関、IT企業などにおいて、LLM開発や科学技術研究、創薬、シミュレーションなどで活用が進んでいる。
- 2023年度以降、**生成AIの登場を契機とした引き合いの急増**が市場成長の大きな要因となっている。今後も研究/教育機関向けをはじめとした大規模な導入が進むとみられ、**好調な市場拡大が予測される**。



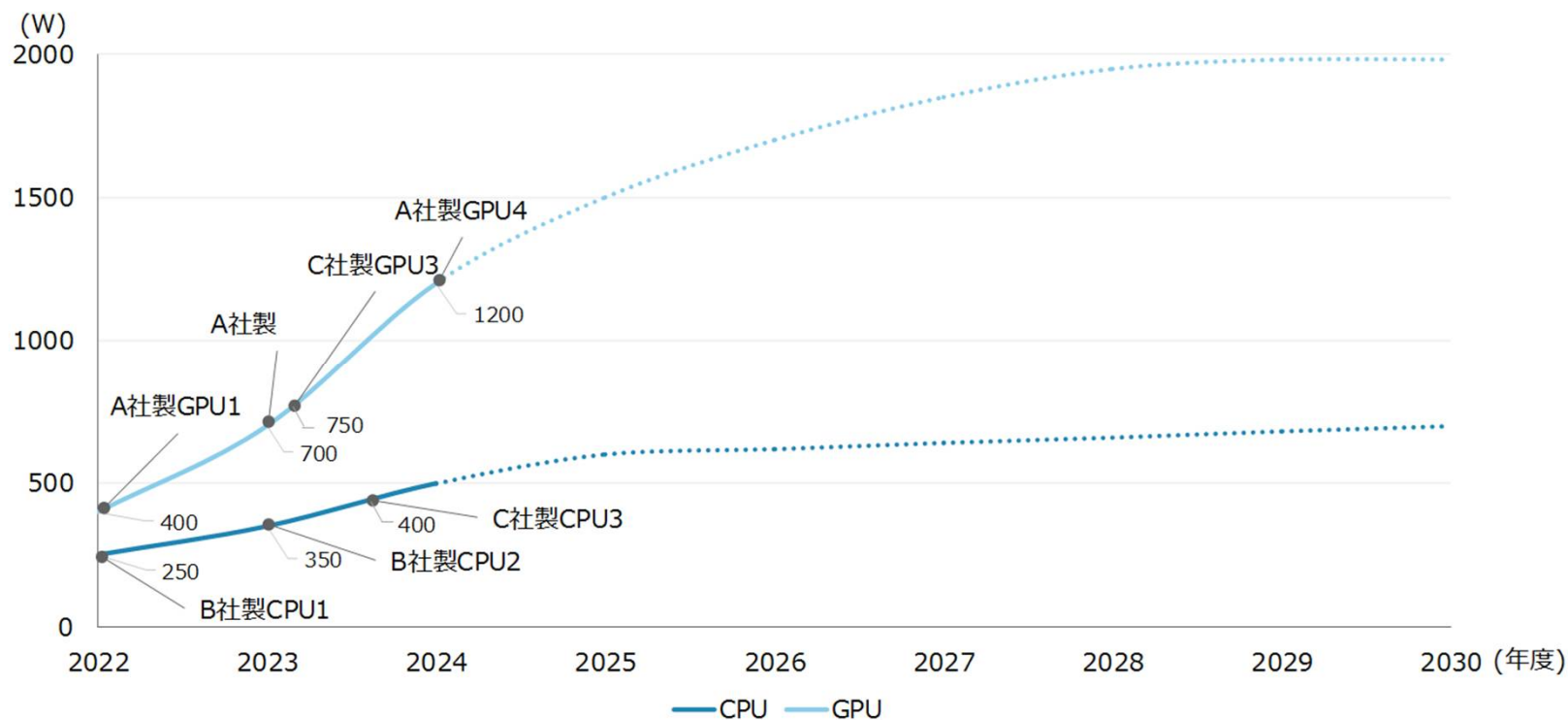
GPUサーバー向けハウジング/GPUクラウド市場規模推移

- 2023年度では、オンプレミスでの稼働が多く、ハウジングやGPUクラウドといったデータセンターでの稼働は合計5,000kWに満たない。
- 中長期的には高発熱対応センターの整備やGPUクラウド需要の拡大などを背景に、データセンターでの稼働が増加する見込み。
- なお商用データセンター全体で提供されるIT電力量(ITロード)は2023年度で1,003,900kW、2027年度には2,490,900kWとなる見通し。
⇒GPUハウジング/クラウドにおける消費電力量(需要量)は、**2027年度で3.8%に留まるとみられる**。



CPU/GPU消費電力量動向

- CPUは350W～400W程度までTDP(Thermal Design Power：熱設計電力)が上昇しており、今後500W～600W程度まで上昇するとみられる。GPUは2023年には700W～750Wまで上昇しているが、高性能化に伴い1,000W～1,200Wまで上昇する見込みである。
- 今後も発熱量は拡大する見込みであり、高発熱機器を設置するデータセンターでは、特に空調/冷却設備の整備が必須となる。



データセンターにおける空調方式の概要

空調方式	概要	メリット	デメリット
空冷方式	<ul style="list-style-type: none"> 空気循環により熱移動を行う、従来型の空調方式。 高発熱ラックを設置する場合は、ラック配置の調整(間引き)を行う場合がある。 	追加費用が不要	20kW/ラックまでの対応となる
液冷方式	<ul style="list-style-type: none"> 冷水をサーバールーム内に引き込み、それをを用いる空調方式。 	高発熱ラックへの対応が可能	追加工事や設備など追加費用が必要
リアドア	<ul style="list-style-type: none"> ラック型空調機により、IT機器から発生した熱を瞬時に除去する局所冷却方式。 	フリークーリングとの併用が可能	コストが高い
DLC	<ul style="list-style-type: none"> CPU/GPUなどの発熱媒体上に設置されたコールドプレート内に直接冷却液を供給する冷却方式。 	50kW/ラック程度まで対応可能	メーカーによって冷媒や温度が異なる
液浸 (单相式/2相式)	<ul style="list-style-type: none"> 絶縁体を持ち、熱交換率の高い液体の中にIT機器本体を浸し、直接熱交換を行う冷却方式。 	省スペース/低コストで設置可能	機器側の運用負担が高い

【用語定義】

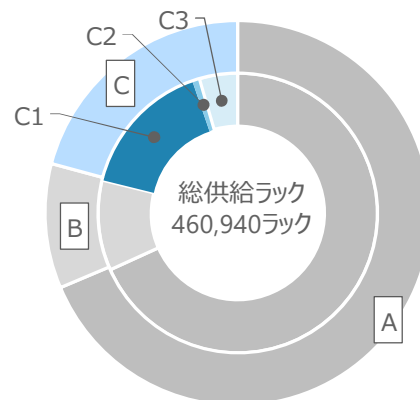
- **熱源設備**：水を冷却し冷温水を発生させる設備。データセンターではチラーやターボ冷凍機が用いられる。
- **CDU**：「Coolant Distribution Unit」の略称で、熱源設備側から供給される冷却すりの温度や排水先をコントロールする設備。
- **フリークーリング**：外気温度の低い中間期(春・秋)や冬季に、チラーや冷凍機ではなく、冷涼な外気を利用した冷却塔で空調用の冷水を発生させるシステム。

冷却方式別国内データセンター供給量(総供給ラック数ベース)

【2023年度】

単位：ラック、%

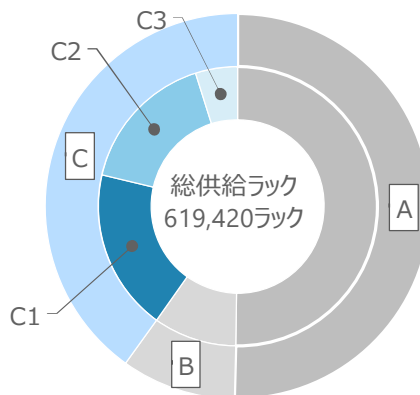
ラック電力スペック	2023年度	
	ラック数	比率
A： ～4kVA (高発熱ラック対応不可)	316,500	68.7
B： 4kVA～6kVA (空冷/間引き)	48,250	10.5
C： 6kVA～ (主に8kVA～10kVA)	96,190	20.9
C1: 空冷/間引き	73,690	76.6
C2: DLC	1,900	2.0
C3: リアドア	20,600	21.4
合計	460,940	100.0



【2027年度】

単位：ラック、%

ラック電力スペック	2027年度	
	ラック数	比率
A： ～4kVA (高発熱ラック対応不可)	311,240	50.2
B： 4kVA～6kVA (空冷/間引き)	59,610	9.6
C： 6kVA～ (主に8kVA～10kVA)	248,570	40.1
C1: 空冷/間引き	116,070	46.7
C2: DLC	102,000	41.0
C3: リアドア	30,500	12.3
合計	619,420	100.0



■ 2023年度

● 液冷対応のセンターは限定的

6kVA～の高発熱ラックのうち、23.4%
高発熱ラックへは空冷/間引きでの対応が中心

■ 2027年度

● DLCを中心に液冷対応センターの整備が進む

高発熱ラックのうちDLCが41.0%、リアドアが12.3%と、
併せて半数以上を占めると予測(53.3%)

今後の高発熱ラック向け冷却対応

■ 空冷方式

一部のセンターにおいては、チラーレス＋外気冷却の空調システム構成で高温帯(35℃程度)で運用するケースもあり、このような空調方式のセンターも多少増加していく模様である。

■ 液冷方式

20kW/1ラック以上となると液冷方式が最適であり、今後もラックあたりの発熱量は増加することから、液冷対応センターの整備は進展するとみられる。

- ▶ リアドア：設置コスト面などから、中長期的には普及しないものとみられる。
- ▶ DLC：多くの新設予定のセンターで採用される見込みで、液冷のデファクトスタンダードになるとみられる。
- ▶ 液浸：実証実験段階であり、実用化には時間を要するものとみられる。

開設予定データセンター GPU冷却対応事例

NO.	開設予定年	IT電力(kW)	採用空調方式	NO.	開設予定年	IT電力(kW)	採用空調方式	NO.	開設予定年	IT電力(kW)	採用空調方式
1	2024	19,200	空冷	15	2025	20,000	未定	29	2026	16,000	未定
2	2024	32,000	空冷	16	2025	19,500	未定	30	2026	17,700	未定
3	2024	20,000	リアドア/DLC	17	2025	72,000	未定	31	2026	36,800	未定
4	2024	30,000	リアドア/DLC	18	2025	36,000	未定	32	2026	48,000	未定
5	2024	36,000	リアドア	19	2025	70,000	未定	33	2026	33,000	未定
6	2024	18,000	空冷	20	2025	16,900	空冷	34	2026	50,000	未定
7	2024	48,500	空冷	21	2025	7,400	未定	35	2026	33,000	未定
8	2024	60,000	空冷	22	2025	6,000	DLC	36	2027	18,000	リアドア
9	2024	48,000	空冷	23	2025	20,000	空冷	37	2027	30,000	未定
10	2024	10,000	空冷	24	2025	48,000	空冷	38	2027	36,000	未定
11	2025	10,000	空冷	25	2026	48,000	未定	39	2027	39,000	未定
12	2025	30,000	DLC	26	2026	20,000	未定	40	2027	35,000	未定
13	2025	40,000	未定	27	2026	60,000	空冷	41	2027	11,000	未定
14	2025	14,000	空冷	28	2026	10,000	未定	42	2028	40,000	未定

データセンターと省エネの関係性

■ ベンチマーク制度

ベンチマーク制度は業種毎で共通の指標(ベンチマーク指標)を定めることで、他事業者との比較による省エネ推進を目的とした制度である。2022年度(令和4年度)よりデータセンター業にも導入されており、事業者はS・A・B・Cの4段階にクラス分けされ評価される。

- ベンチマーク指標：PUE(Power Usage Effectiveness)=データセンター施設全体のエネルギー使用率 ÷ IT機器のエネルギー使用量
- Sクラス：**PUE=1.4以下** 優良事業者として経産省HPIに事業者名や連続達成年数を表示



PUE改善が課題に



■ PUE改善に向けて

PUE改善にあたり、電力使用量がIT機器に次いで大きい**空調設備関連の消費電力削減が必要**となる。

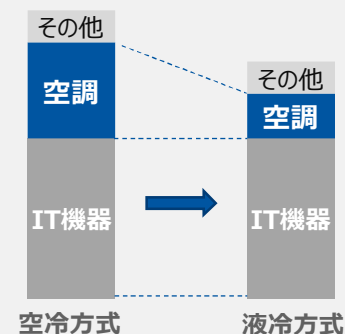
- 空調の消費電力削減：サーバールーム内温度=25℃±2℃を保ち、過冷却を防ぎながら、安定的で最適な空調機の稼働が必須

■ 高発熱ラック対応

生成AI関連市場の需要拡大に伴い、GPUサーバーを搭載した高発熱ラックは今後も増加する。

こうした高発熱ラック向けには、空冷方式よりも熱交換率の高い**液冷方式の採用も消費電力削減に有効**であり、省エネ促進の観点からもデータセンター空調設備の液冷対応は進展するものと期待される。

空調方式別消費電力イメージ



JEITA

JEITA

**Japan Electronics and Information
Technology Industries Association**

