

Google Next 2024 – Day 1 Recap
 Courtesy of @mikelor
<https://www.linkedin.com/in/mlorenzo>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
Compute	A3 Mega VMs	A3 Mega VMs feature Nvidia H100 Tensor Core GPUs, each with 80 billion transistors on a single chip.	General Availability	Allows training of more complex ML models at a fast speed to build large language models (LLMs), generative AI, and diffusion models.
	Nvidia GB200 NVL72	A multi-node rack-scale system that will be combined with Google Cloud's fourth generation of advanced liquid-cooling systems.	Early 2025	NVIDIA is a key player in GPU based computing systems. All three clouds have now announced support for the latest NVIDIA GB200/H200 platform.
	TPU v5p	Google's most powerful and scalable TPU to date. Introducing Cloud TPU v5p and AI Hypercomputer	General Availability	Google's TPU v5p is a powerful chip for AI workloads,
	New Axion Chip	Google's first chip for the data center. 30% better performance than the fastest general-purpose Arm-based instances available in the cloud today. Google Announces Homegrown ARM-based CPUs	Coming Soon	A custom-designed accelerator focused on accelerating specific workloads, particularly machine learning tasks.

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenco>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
Storage	AI Optimized Storage	Cloud Storage FUSE: This makes high-performance, low-cost cloud storage accessible to complex AI/ML applications.	Preview	By reducing data access latency and improving data throughput, these advancements can significantly accelerate AI/ML training and inference processes.
		Filestore Optimization: This optimizes the Filestore network file system for AI/ML models.		
		Hyperdisk ML, a new block storage service optimized for AI inference/serving workloads,		
Vertex AI Models	Gemini 1.5 Pro	Brings the world's largest context window (1M tokens) to developers. Enough to process 1Hr of video, 11 Hrs of audio, 30K lines of code, or 700K words.	Preview	A larger Context Window (Tokens) allows the model to reason over much larger amounts of data.
	Claude 3	Anthropic's latest state-of-the-art model. Claude 3 Sonnet and Claude 3 Haiku available now. Claude 3 Opus in the coming weeks.	General Availability	Claude is a large language model (LLM) built by Anthropic. It's trained to be a helpful assistant in a conversational tone.
	CodeGemma	A new fine-tuned version of Gemma designed for code generation and code assistance Gemma Family Expands with Models Tailored for Developers	General Availability	Code Gemma is seen as a competitor to existing code completion and generation tools like Github's Copilot. With its open-source nature and focus on multiple languages,

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenzo>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
Vertex AI Tools	Agent Builder	New. Brings together foundation models, Google Search, and other developer tools, to build and deploy agents. A no-code agent builder console with powerful grounding, orchestration and augmentation capabilities. Build generative AI experiences with Vertex AI Agent Builder	Generally Available	Agent Builder is an expansion and rebranding of Vertex AI Search and Conversation. You can create AI agents using natural language or a code-first approach.
	Imagen 2.0	Updated Version of text-to-image generation technology. New text-to-live image capability allows marketing and creative teams to generate animated images. Advanced photo editing features, including inpainting and outpainting, Added Digital Watermarking for AI-generated images.	Preview	New text-to-live image capability creates 4 second videos for creating engaging content.
	Prompt Management	Collaborate on built-in prompts with notes and status, track changes over time, and compare the quality of responses from different prompts.	Preview	Prompt management acts as a bridge between you and the generative AI model. By effectively crafting prompts, you unlock the full potential of the model and achieve the specific results you desire.

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenzo>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
	Automatic side-by-side (AutoSxS)	Provides explanations of why one response outperforms another and certainty scores, which helps users understand the accuracy of the evaluation.	Generally Available	Streamlines the often complex and time-consuming process of training, comparing, and selecting machine learning models.
	Rapid Evaluation feature	Helps customers quickly evaluate models on smaller data sets when iterating on prompt design.	Preview	Offers advantages for streamlining the machine learning development process by enabling faster model assessment and improved selection.
Vids	New Product	<p>New AI-powered video creation app for work is your video, writing, production, and editing assistant, all in one. Vids will sit alongside other Workspace apps like Docs, Sheets, and Slides.</p> <p>5 Workspace updates from Google Cloud Next 2024</p>	Private Preview	A faster way to create engaging videos with generative AI, and a conversational interface.
Meetings and Messaging	New Add-On	With "take notes for me", chat summarization, and real-time translation in 69 languages (equal to 4,600 language pairs), this collaboration tool will only cost \$10 per user, per month.	General Availability (?)	Individual productivity gains.
Security	New Add-On	Workspace admins can now automatically classify and protect sensitive files and data using privacy-preserving AI models and Data Loss Prevention controls trained for their	General Availability (?)	Leverages AI for enhanced security within workspaces by automating threat detection, enabling customizable data protection, and improving efficiency.

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenzo>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
		organization. The AI Security add-on is available for \$10 per user, per month		
Big Query	Gemini in Big Query	Uses AI to help your data teams with data preparation, discovery, analysis and governance.	Preview	By leveraging AI and offering a more intuitive experience, it can empower data professionals to extract valuable insights from their data more efficiently.
	Data Canvas	Provides a new notebook-like experience with natural language and embedded visualizations	Preview	By offering a visual interface and reducing reliance on code, it can empower a wider range of users to explore and gain valuable insights from data stored in BigQuery.
Databases	Gemini in Databases	Makes it easy for you to migrate data safely and securely from legacy systems, for example, converting your database to a modern cloud database like AlloyDB	Preview	Leverages AI to make database management more efficient, accessible, and secure. It empowers developers to focus on innovation while AI handles the routine tasks.
	Vector Indexing	New capabilities using vector indexing directly in AlloyDB and BigQuery allow you to leverage AI over your data where it is stored, enabling real-time and accurate responses.	Preview	Enables efficient searches based on meaning and relationships, it opens doors for more powerful search functionalities, improved machine learning applications, and support for emerging AI workloads.
Code Assist	Gemini 1.5 Pro in Code Assist	Allows for 1 million token context window, revolutionizing coding for even the largest projects. Delivers even more accurate code suggestions, deeper insights, and streamlined workflows.	Private Preview	The larger token context window allows for analyzing large system codebases.

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenco>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
Cloud Assist	Gemini in Cloud Assist	Design, secure, operate, troubleshoot, and optimize app performance and costs.	Preview	By leveraging AI to automate tasks, improve efficiency, and enhance security, it has the potential to accelerate innovation and make cloud computing more accessible and beneficial for a wider range of users.
Threat Intelligence	Gemini in Threat Intelligence	Uses natural language to deliver deep insight about threat actor behavior. Gemini's larger context window allows for analysis of the interactions between modules, providing new insight into code's true intent.	Preview	Leverages artificial intelligence (AI) and machine learning (ML) to bolster an organization's security posture by providing advanced threat detection and mitigation capabilities.
Security Operations	Gemini in Security Operations	A new assisted investigation feature converts natural language to detections, summarizes event data, recommends actions to take, and navigates users through the platform via conversational chat.	Preview	By leveraging AI and natural language processing, it has the potential to significantly improve security team efficiency, threat analysis capabilities, and overall incident response effectiveness.
Goldman Sachs	David Solomon Chairman & CEO		Video Segment	
Mercedes Benz	Ola Kallenius		Video Segment	
Walmart	Suresh Kumar EVP, Global Chief Technology Officer		Video Segment	
Uber	Dara Khosrowshahi CEO		Video Segment	
Wayfair	Fiona Tan		Video Segment	

Google Next 2024 – Day 1 Recap

Courtesy of @mikelor

<https://www.linkedin.com/in/mlorenzo>

Product	Feature / Announcement	Description and Link to More Information	Availability	What's Interesting / Why It Matters
		Chief Technology Officer		
Palo Alto Networks	Nikesh Arora CEO		Video Segment	