



Desbloqueando el Potencial de la IA Generativa con AWS

Sergio Garcés Vitale

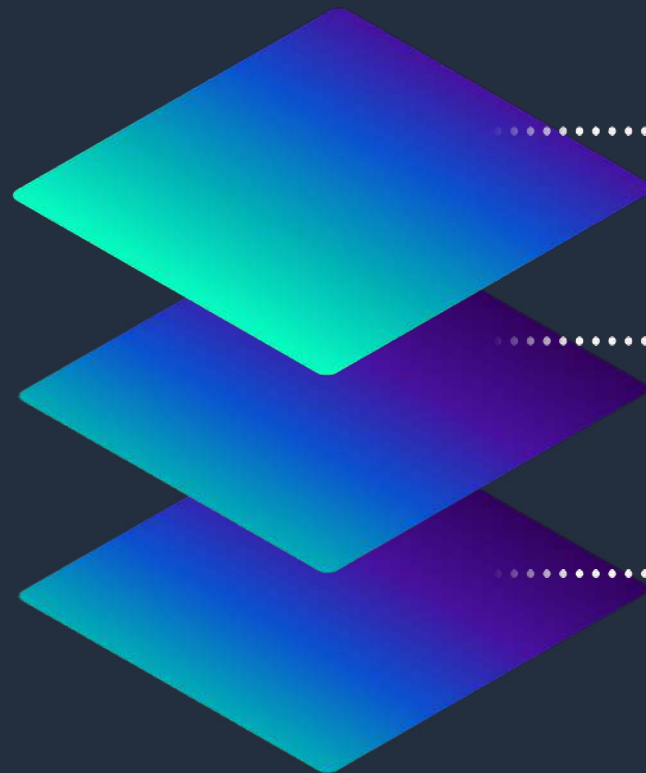
Senior Solutions Architect
AWS



Sergio Garcés Vitale
Senior Enterprise Solutions Architect en
Amazon Web Services (AWS)




The Generative AI Stack












APLICACIONES QUE USAN FMs

-  Amazon Q
-  Amazon Q in Amazon QuickSight
-  Amazon Q in Amazon Connect
-  Amazon CodeWhisperer

HERRAMIENTAS QUE CONSTRUYEN CON FMs Y LLMs

-  Amazon Bedrock
- Guardrails
- Agents
- Customization capabilities

INFRAESTRUCTURA PARA TRAINING E INFERENCIA

-  GPUs
-  Trainium
-  Inferentia
-  SageMaker
-  UltraClusters
-  EFA
-  EC2 Capacity Blocks
-  Nitro
-  Neuron

Amazon Bedrock

Elegir no es un privilegio,
es un acto de responsabilidad

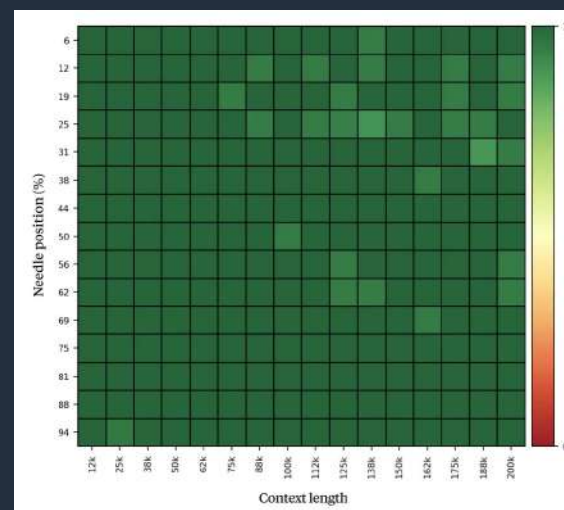


Amazon Bedrock

Anthropic Claude 3

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge <i>MMLU</i>	86.8% 5-shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	—	—
Grade school math <i>GSM8K</i>	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot	94.4% Majl@32	86.5% Majl@32
Math problem-solving <i>MATH</i>	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math <i>MGSM</i>	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	—	79.0% 8-shot	63.5% 8-shot
Code <i>HumanEval</i>	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text <i>DROP, FI score</i>	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed evaluations <i>BIG-Bench-Hard</i>	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Knowledge Q&A <i>ARC-Challenge</i>	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	—	—
Common Knowledge <i>HellasSwag</i>	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84.7% 10-shot

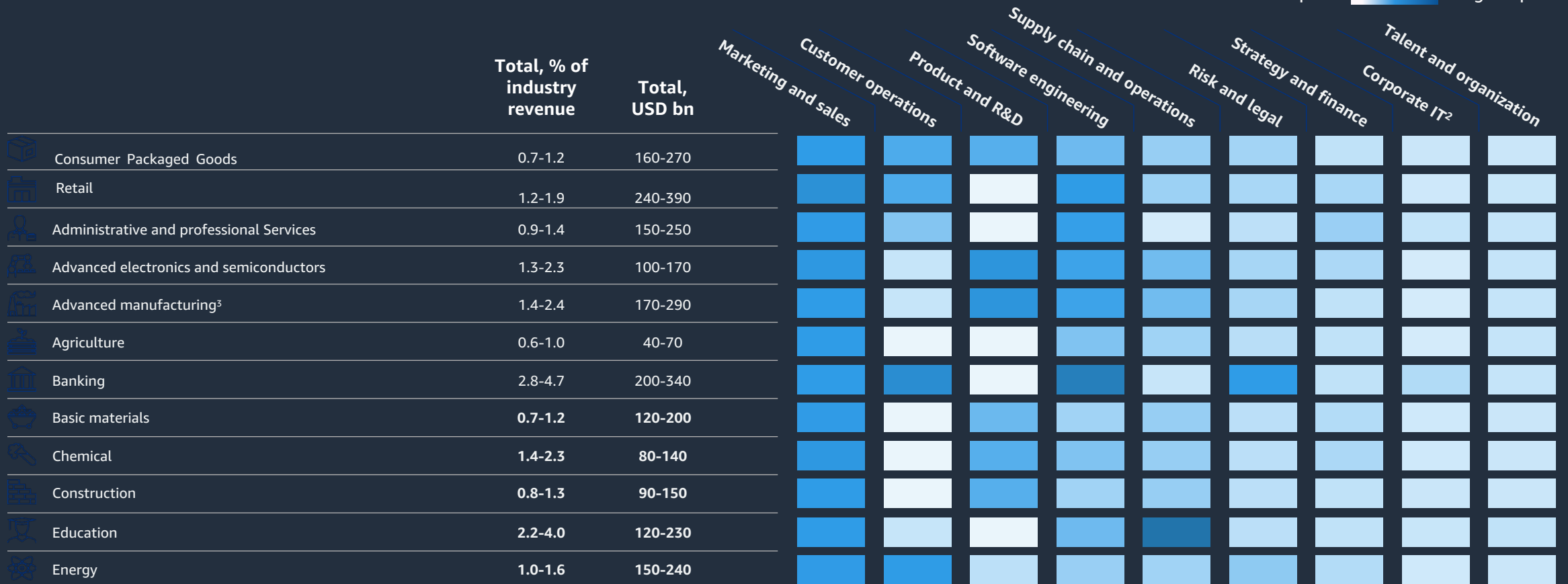
	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4V	Gemini 1.0 Ultra	Gemini 1.0 Pro
Math & reasoning <i>MMMU (val)</i>	59.4%	53.1%	50.2%	56.8%	59.4%	47.9%
Document visual Q&A <i>ANLS score, test</i>	89.3%	89.5%	88.8%	88.4%	90.9%	88.1%
Math <i>MathVista (testmini)</i>	50.5% CoT	47.9% CoT	46.4% CoT	49.9%	53.0%	45.2%
Science diagrams <i>AIZD, test</i>	88.1%	88.7%	86.7%	78.2%	79.5%	73.9%
Chart Q&A <i>Relaxed accuracy (test)</i>	80.8% 0-shot CoT	81.1% 0-shot CoT	81.7% 0-shot CoT	78.5% 4-shot CoT	80.8%	74.1%



Generative AI use cases will have different impact on business functions across industries – however, four functions prevail across industries

Estimated global, annual Generative AI productivity impact, by business functions and industry¹

Low impact  High impact



Source: McKinsey & Company (2023): The economic potential of Generative AI: The next productivity frontier

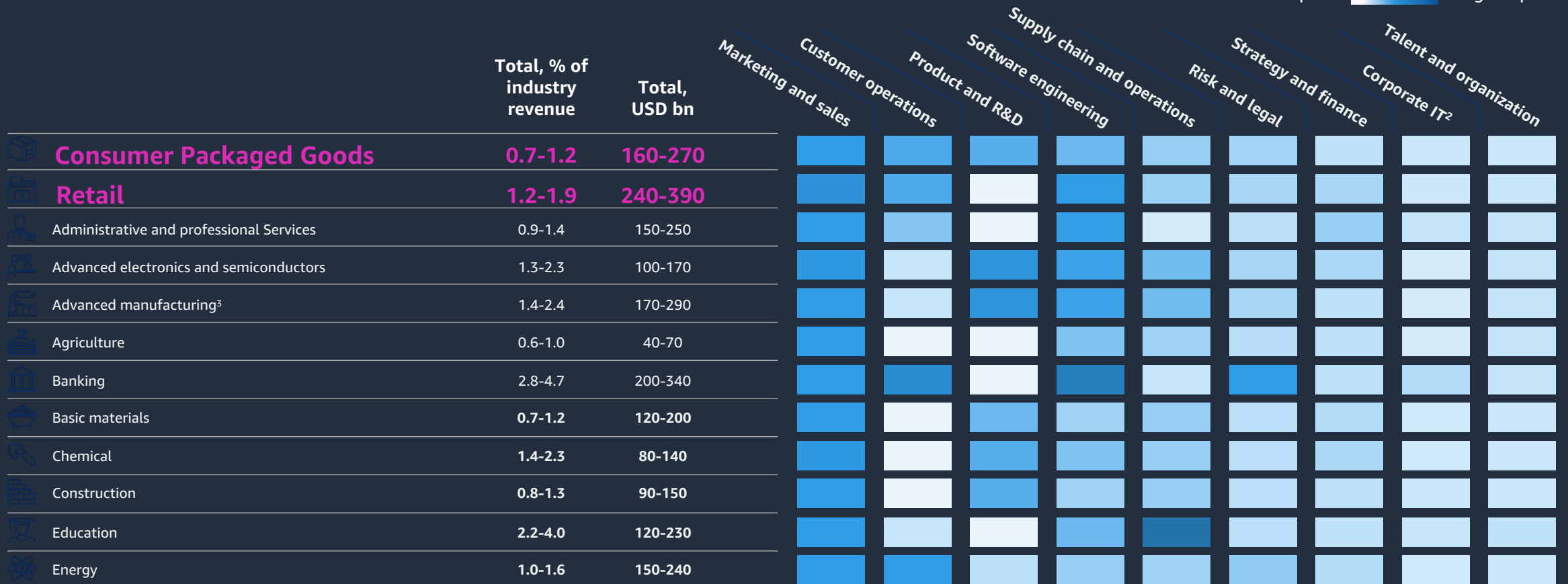
1. Excludes implementation costs (e.g., training, licenses) | 2. Excluding software engineering | 3. Includes aerospace, defense, and auto manufacturing



Generative AI use cases will have different impact on business functions across industries – however, four functions prevail across industries

Estimated global, annual Generative AI productivity impact, by business functions and industry¹

Low impact  High impact



Source: McKinsey & Company (2023): The economic potential of Generative AI: The next productivity frontier

1. Excludes implementation costs (e.g., training, licenses) | 2. Excluding software engineering | 3. Includes aerospace, defense, and auto manufacturing



IA Generativa en acción

... más allá del PPT



Conoce a tu
cliente



Identifica tu
producto



Diseña tu
campana

DEMO





**“A journey of thousand miles
begins with a single step”**

Lao Tzu

Founder of philosophical and religious Taoism





Gracias!



Sergio Garcés Vitale

Senior Enterprise Solutions Architect en
Amazon Web Services (AWS)

