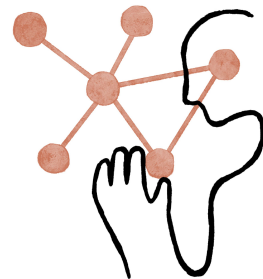


ANTHROPIC

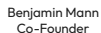
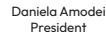
March 13, 2024



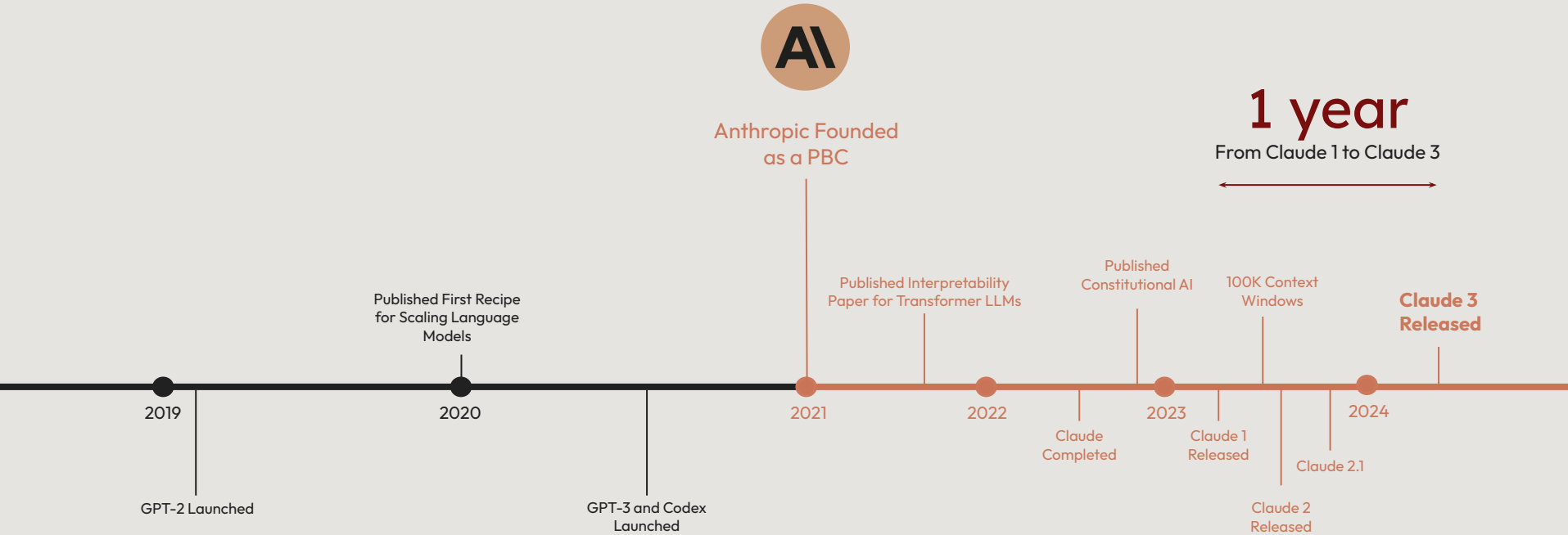
Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*
Jared Kaplan	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam
Girish Sastry	Amanda Askell	Sandhini Agarwal	Ariel Herbert-Vass
Gretchen Krueger	Tom Henighan	Rewon Child	Aditya Ramesh
Daniel M. Ziegler	Jeffrey Wu	Clemens Winter	Christopher Hesse
Benjamin Chess	Sam McCandlish	Mark Chen	Eric Sigler
Mateusz Litwin	Scott Gray	Jack Clark	Christopher Berner
Alec Radford	Ilya Sutskever	Dario Amodei	

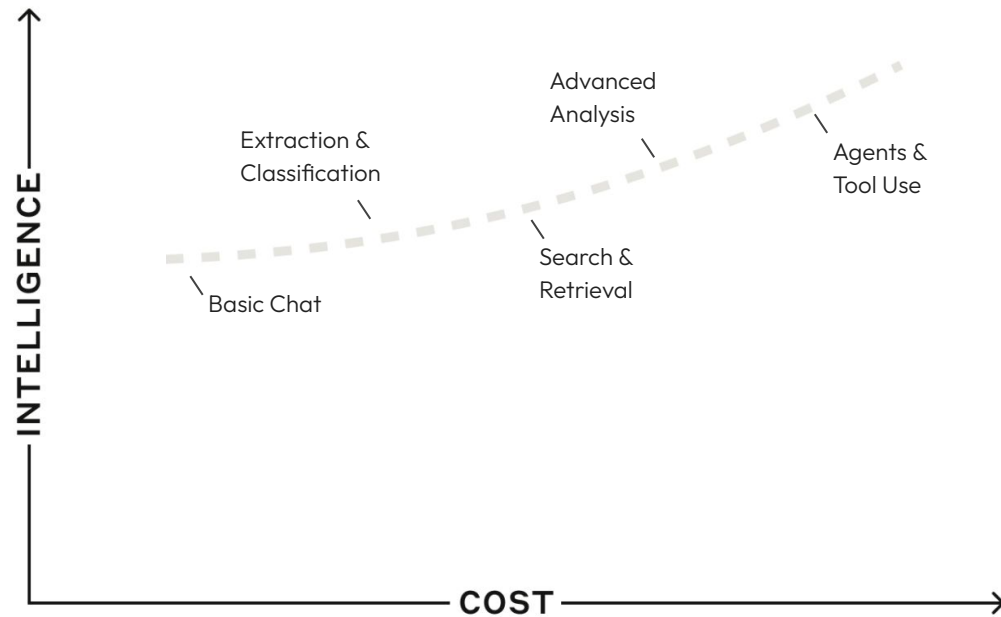
Co-Founders are the authors of the defining GPT-3 paper



Our breakthroughs continue to define the AI industry



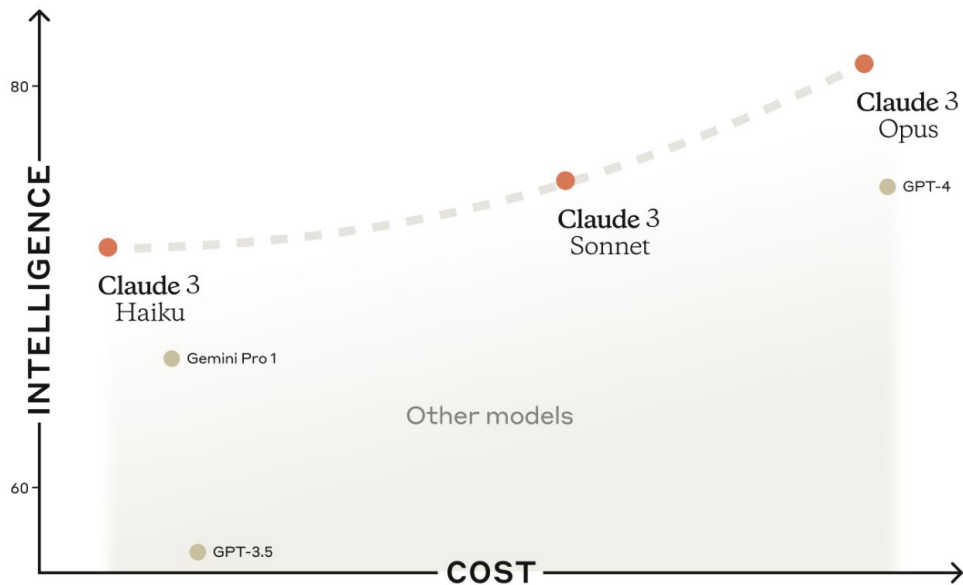
There's no “one size fits all” model for enterprise AI



**Image is for illustrative purposes only and not to scale*

Claude 3

Leading the frontier of **speed**, **intelligence**, and **cost-efficiency** for enterprise AI



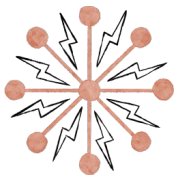
*Intelligence score (percentage) is an average of top published benchmarks for each model

A new standard for model intelligence

	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% Maj1@32	86.5% Maj1@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, F1 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>HellaSwag</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

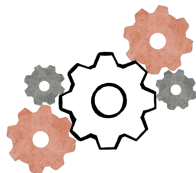
Purpose-built for enterprise AI requirements

Faster



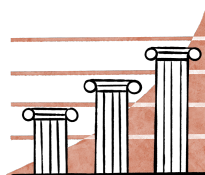
Faster models available in each intelligence class

More steerable



Better results out-of-the-box with less prompt optimization and fewer refusals

More accurate & trustworthy



Twice as accurate as Claude 2.1 on difficult, open-ended questions

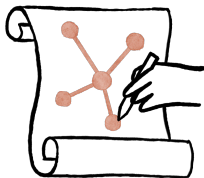
Vision



The fastest vision model with comparable quality to other state-of-the-art models

Constitutional AI allows us to build safer AI at scale

Constitutional Principles



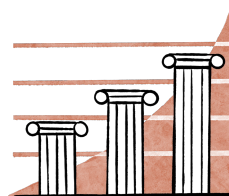
We codify a set of principles to reduce harmful behavior

Efficient AI Generated Datasets



This technique does not require time-intensive human feedback data sets, but rather more efficient AI-generated datasets

Improved and Aligned Outputs



The output of the system is more honest, helpful, and harmless

Listed as one of the

“ The 3 Most Important AI Innovations of 2023 ”

-TIME Magazine, December 2023

Claude 3 Opus

The **best model in the world** for complex, cutting-edge use cases

Core features

- The **world's most intelligent model** available to date
- **State-of-the-art scores** for reasoning, knowledge, math, multilingual performance, and more
- **Industry-leading benchmarks** for business multimodal applications

Top skills

- Advanced logical reasoning
- Agentic behavior
- Tool use
- Graduate-level math
- Data analysis
- Interactive & complex coding

Top use cases

- **Accelerate R&D** tasks like literature review, hypothesis generation, and more
- **End-to-end automation** of complex tasks with agents & tool use
- **Expert-level virtual analyst** to accelerate decision-making
- **Advanced, interactive coding** to solve complex problems

Claude 3 Sonnet

The **best combination of intelligence and speed** for deployments at scale

Core features

- **2x faster** than Claude 2.1
- **Image-to-text vision** trained for business applications
- **Less than half the cost** of comparable leading models, with similar levels of intelligence

Top skills

- Search & Retrieval, RAG
- Image analysis
- Coding
- Content Generation
- Summarization
- Advanced translation

Top use cases

- **Data analysis & insights** over vast amounts of enterprise knowledge
- **Forecasting** for financial and market models
- **Personalized customer analytics & marketing**
- **Boost productivity** for across day-to-day knowledge work

Claude 3 Haiku

The **fastest, most cost-effective model** for high-volume, rapid use cases

Core features

- **Much faster** than other models in its class
- **Lightning-fast vision** capabilities for image-to-text
- **Cheaper, faster, and more intelligent** than comparable lightweight, low-latency models

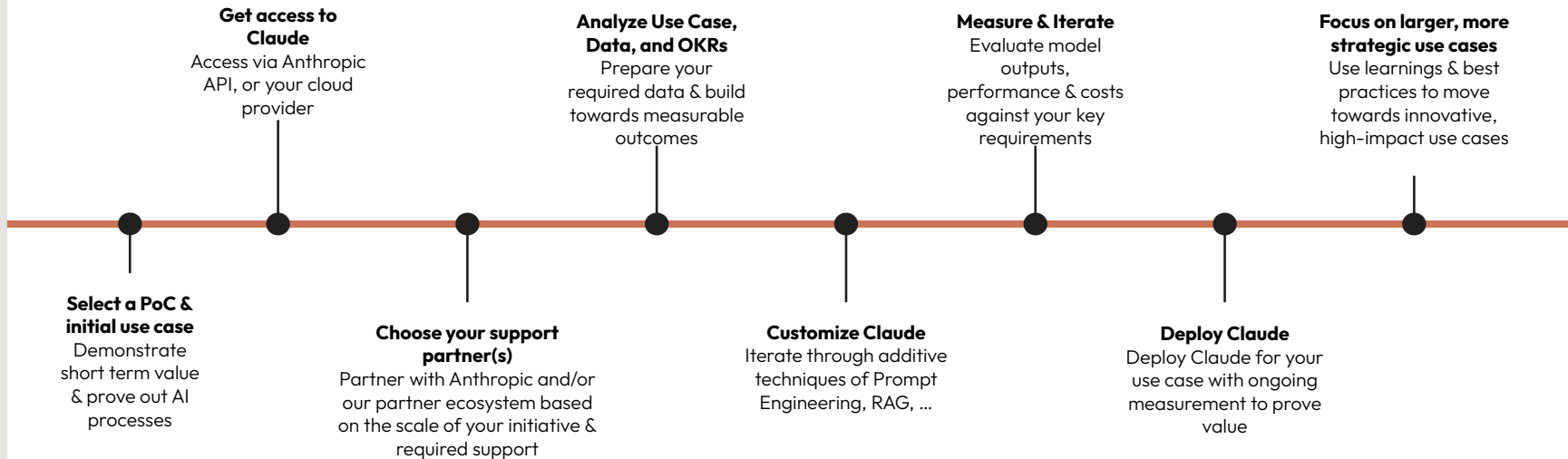
Top skills

- Dialogue
- Data extraction & classification
- Search & Retrieval
- Summarization
- Image analysis
- Translation

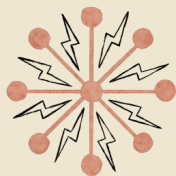
Top use cases

- **Live, interactive chat bots** with multilingual support
- **Real-time content moderation** at scale
- **Data extraction** from both text and images
- **Fast, low-cost translation** at scale

Your path from ideation to successful deployment



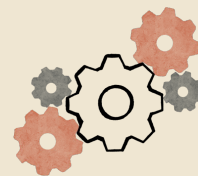
Start today to prepare for a **transformative** future



We expect frontier models will **rapidly transform industries** with new capabilities in 6-18 months.



Our customers are rapidly building on use cases to **maximize their readiness** for future breakthroughs.



Customers who wait for new breakthroughs & capabilities may already be **4-8 months behind**.

AI