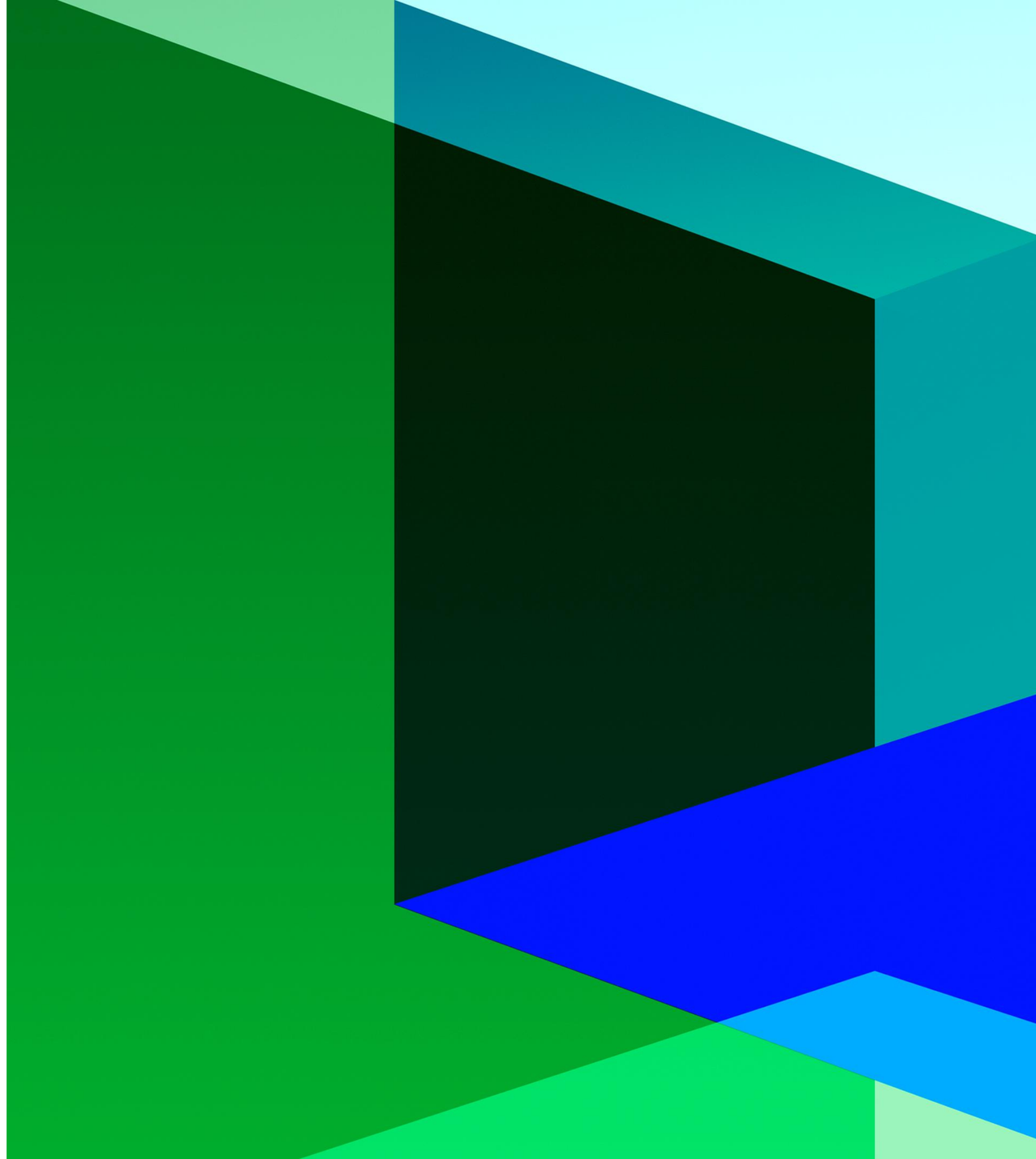
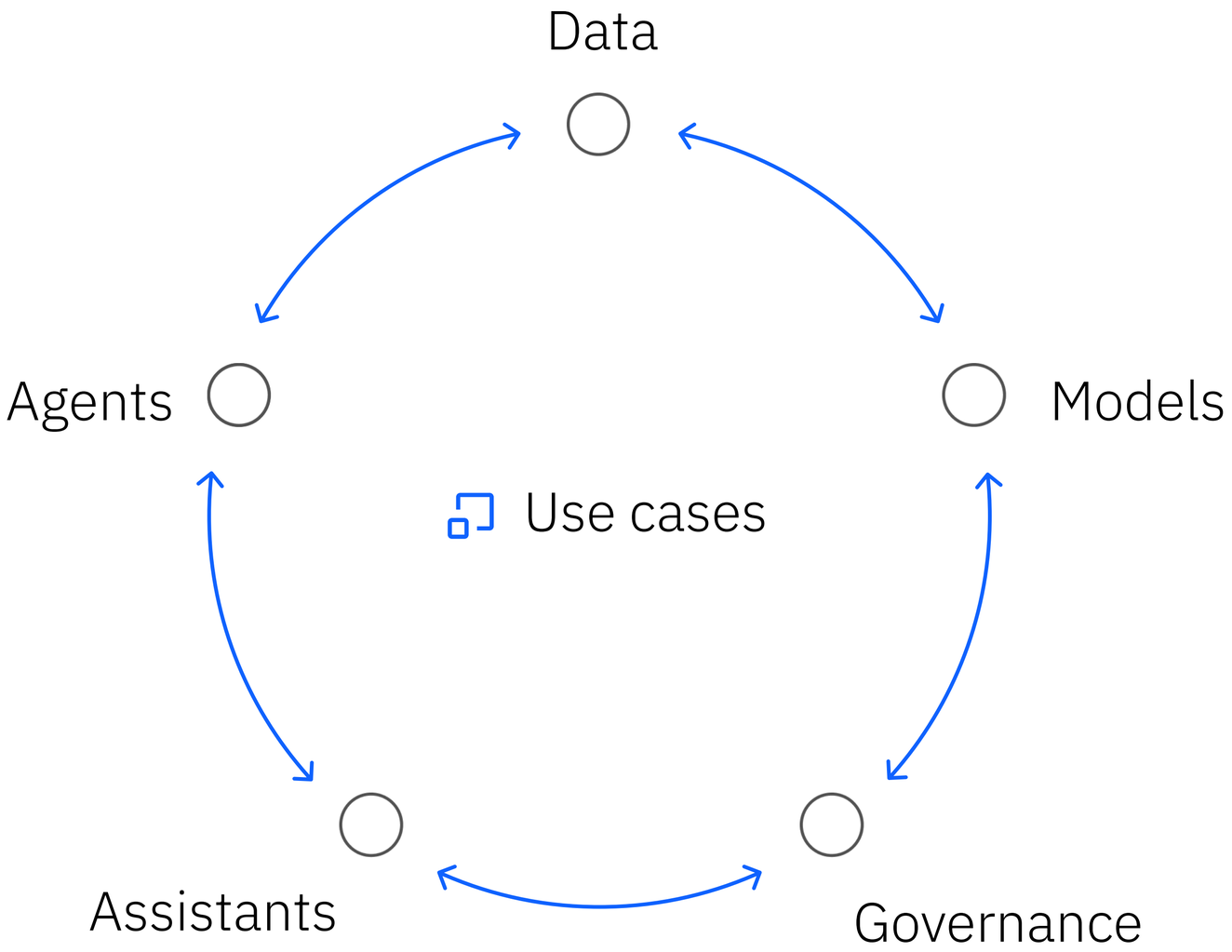
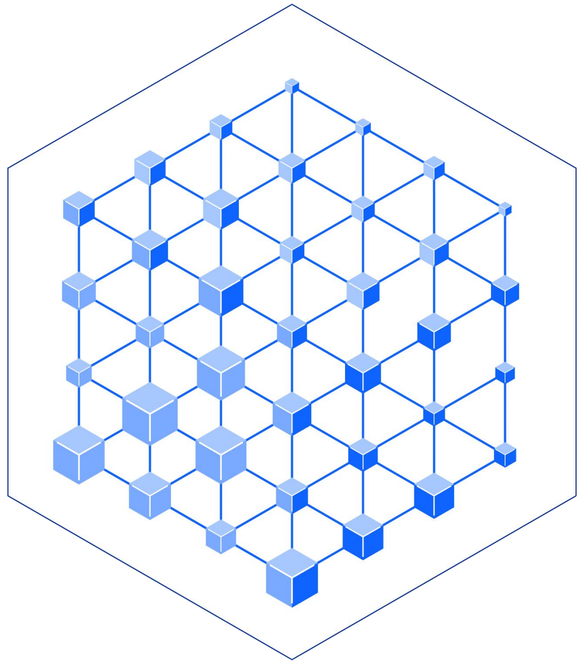
Introducing Granite



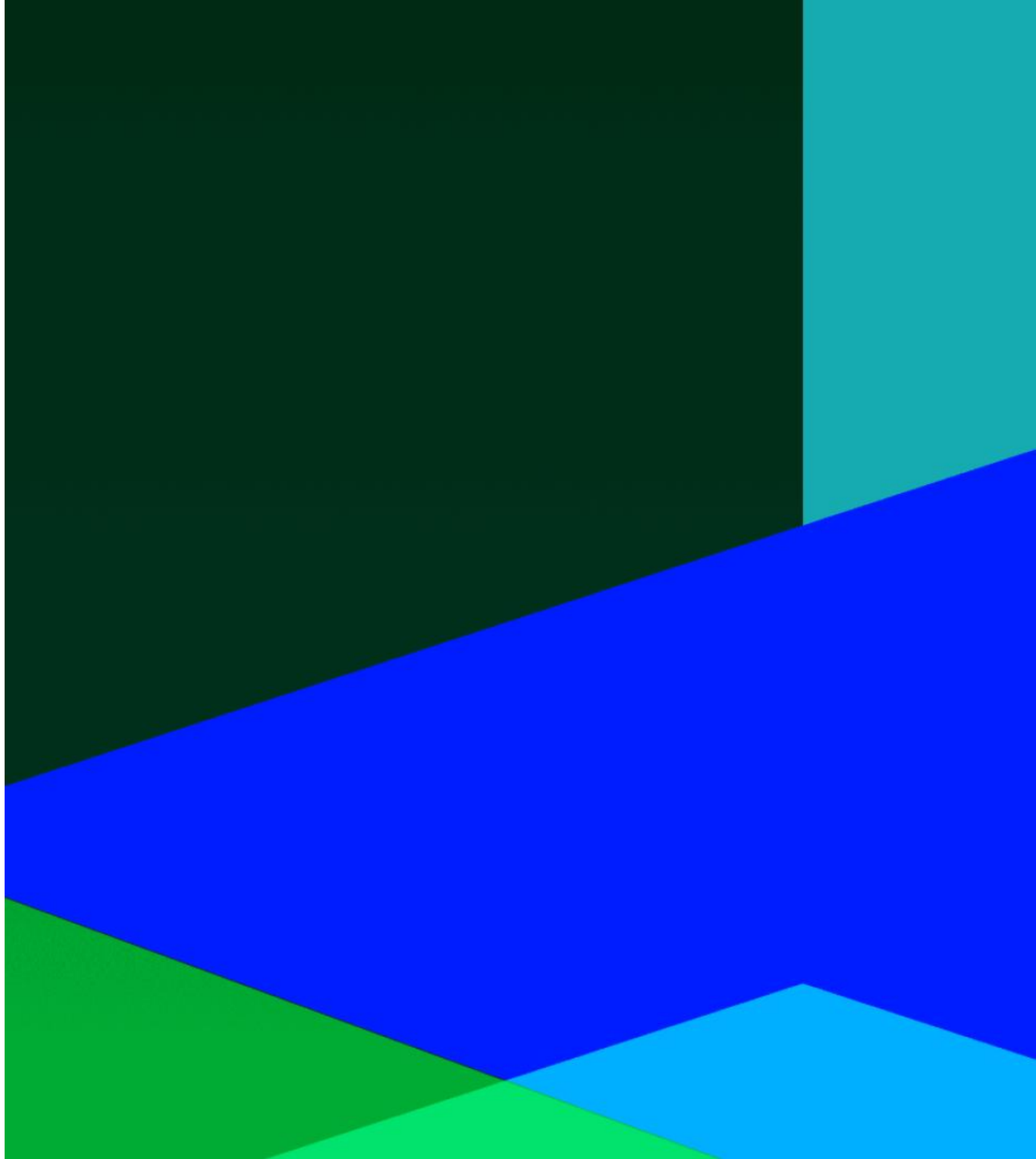
AI building blocks



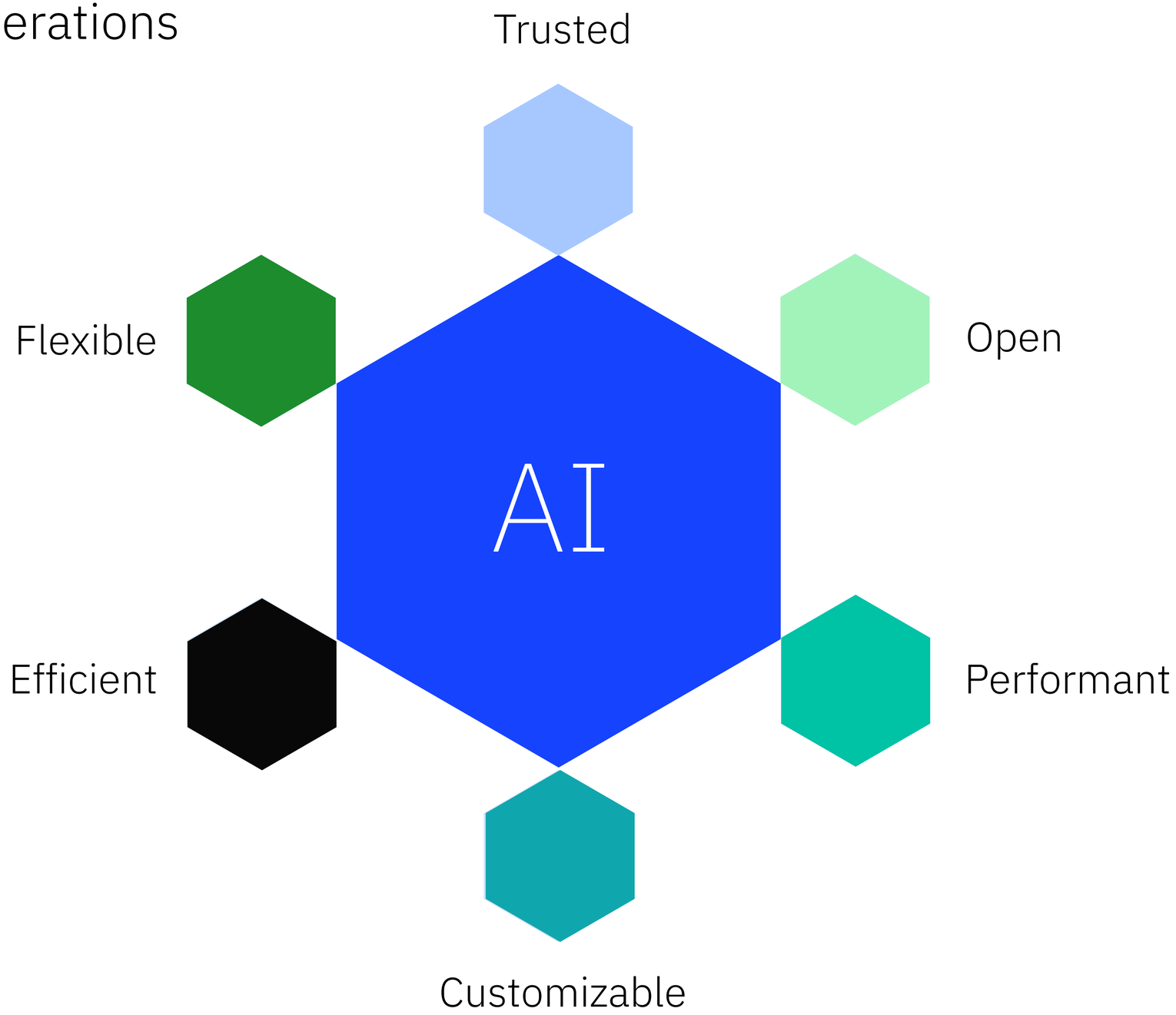
Models



Are your models tailored for business and optimized to scale AI?

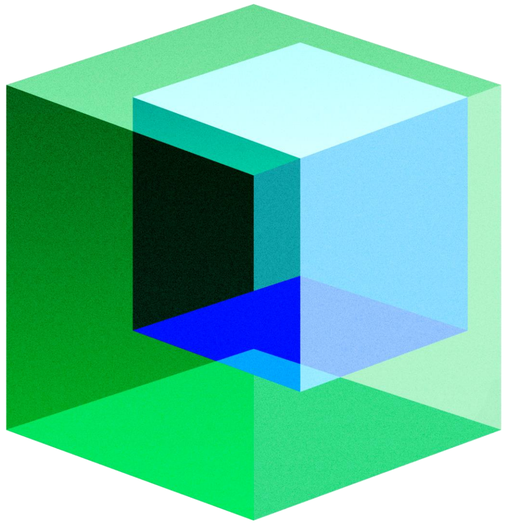


Enterprise Considerations for AI Models



IBM Granite

A family of open, performant and trusted AI models to accelerate enterprise AI adoption



Open

- Open sourced under Apache 2.0
- Transparency of data and training methods
- Customize with your data

Performant

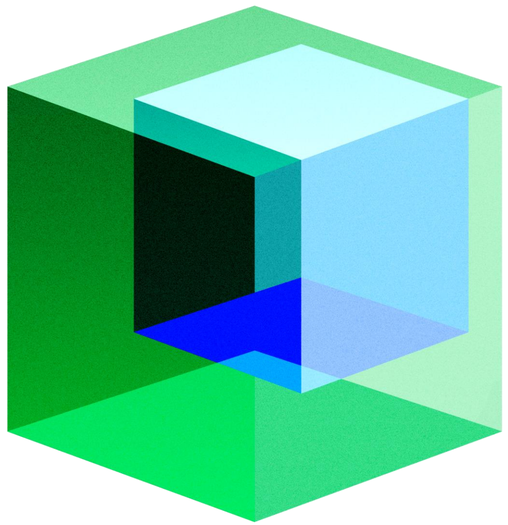
- Diverse range of fit-for-purpose models
- Designed for scalability
- Reasoning capabilities which uniquely maintain general performance
- Vision capabilities optimized for enterprise document understanding

Trusted

- IP indemnification
- Responsible and safe AI
- Guardrails to secure data and mitigate risks
- Reasoning approach which uniquely preserves model safety

IBM Granite

A family of open, performant and trusted AI models to accelerate enterprise AI adoption



Granite 3.0
October 2024

- 6 models trained on 12T tokens across 12 languages and 116 programming languages

Granite 3.1
December 2024

- Extended context length to 128k
- New embedding models
- New time series models

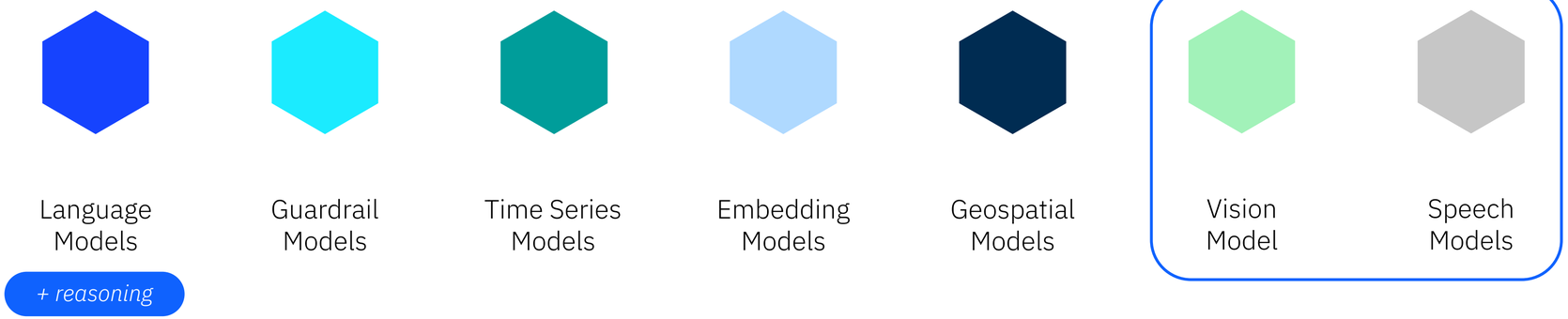
Granite 3.2
February 2025

- Reasoning capabilities
- New vision model

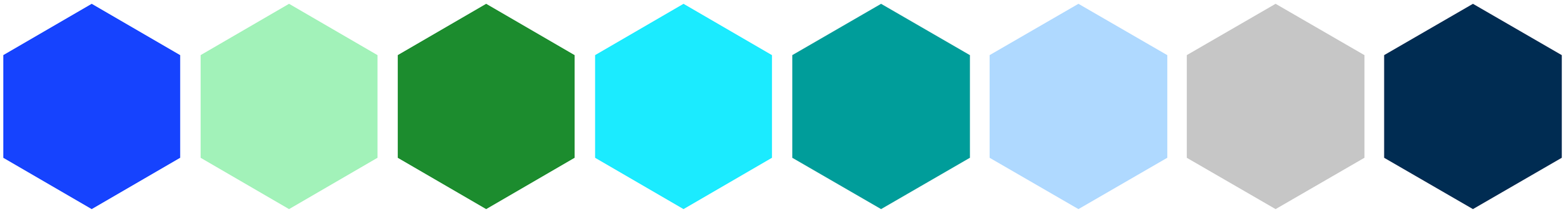
Granite 3.3
April 2025

- New speech model
- Language model improvements
- FIM

Granite family of models



Granite Models



Reasoning capabilities which preserve general performance and safety

Multimodal capabilities optimized for enterprise document understanding

Over 12T tokens training data across 12 languages and 116 programming languages

Open source under Apache 2.0, offering broad commercial usage

Top enterprise generative tasks powered by Granite

Power key generative AI tasks to drive your enterprise use cases

Agentic workflows

Automate tasks, streamline processes, and enhance operational efficiency with AI agents for business

Example: Autonomous HR agents to support Employee Support, Talent Acquisition, and Onboarding

Language-based tasks

Retrieval augmented generation (RAG), summarization, content generation, insight extraction, and classification based on documents or dynamic content

Example: Building a Q&A resource from a broad knowledge base, providing customer service assistance

Code

Optimize the software development lifecycle with code generative tasks, including code generation, code explanation, and code editing

Example: AI-generated code recommendations , IT application modernization from COBOL to Java

Time series

Time-series forecasting to easily analyze current data to make predictions and help make informed decisions

Example: Predicting future customer demand for a given product and period, using historical sales and other data sources

Geospatial

Uncover patterns and trends in geo data

Example: NASA and IBM teamed up to create an AI Foundation Model for Earth Observations using large-scale satellite and remote sensing data

Safety

Safeguard AI with models ensuring enterprise data security and mitigate risks across a variety of user prompts and LLM response

Example: AI compliance with regulatory requirements in financial services, healthcare, and government.

Cooking with Granite

https://ibm.github.io/granite-workshop

https://github.com/IBM/granite-workshop

https://ibm.biz/granite-community