Red Hat Vision for AI and Open Source

Chris Wright

Chief Technology Officer and Senior Vice President,
Global Engineering
Red Hat

Red Hat

Where do we go next?

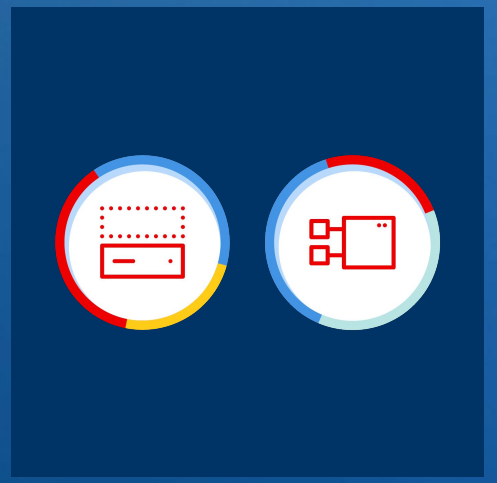
What’s challenging your
IT operations?

What’s driving
your innovation?

Red Hat enables open hybrid cloud

Any cloud

Secure automated infrastructure



Any application

Team collaboration

Open hybrid cloud is evolving

Mission-critical automation

Developer productivity

Trusted security

Edge

AI

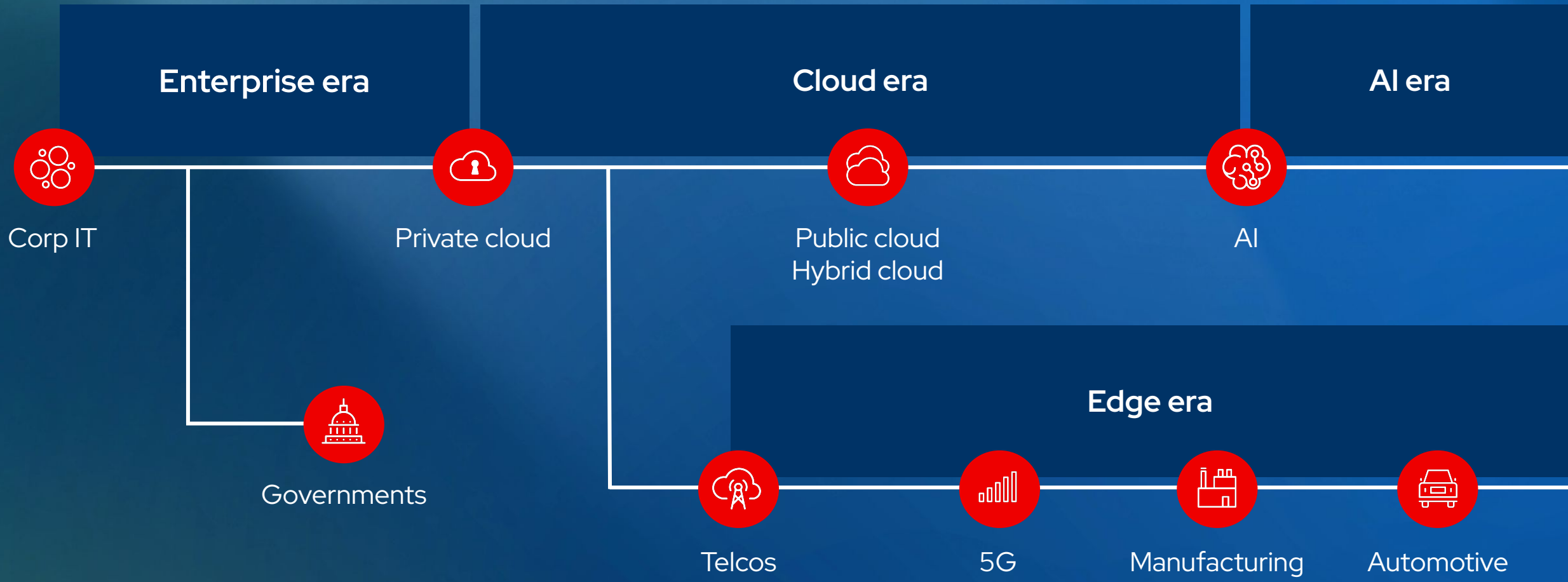
Open Hybrid Cloud

open source and AI

Generative
AI era

You are here

Every previous era has grown and expanded with “open”

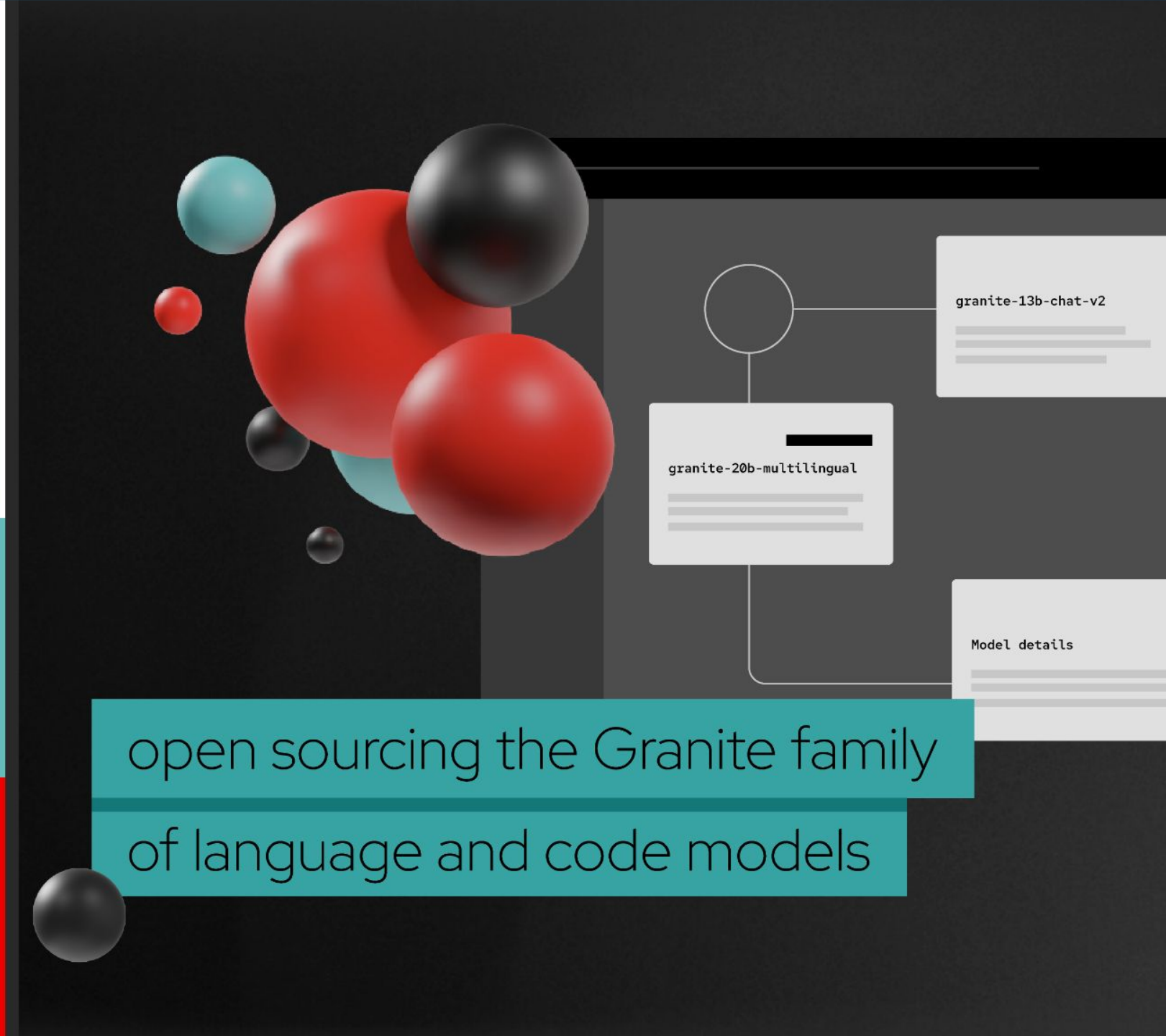


We see a future where open source plays a huge role in AI

- Open source models

Red Hat

IBM Research



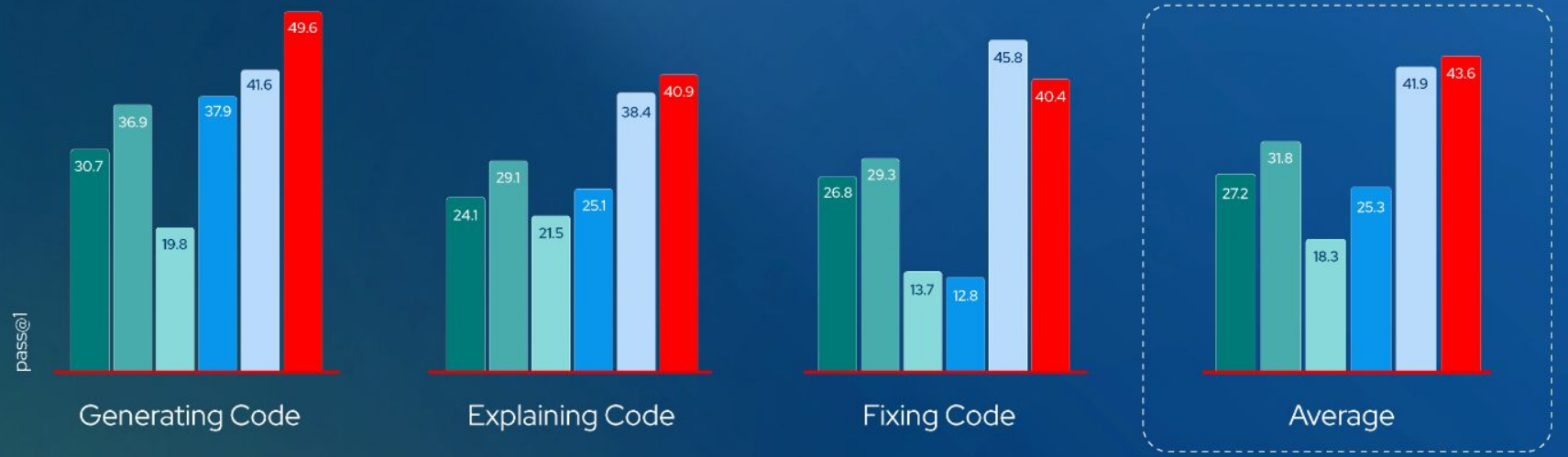
open sourcing the Granite family of language and code models

Big results from **smaller models**

Big results using **data** provided by **experts across your organization**

And big results for much **lower cost**

Code assistance model performance



HumanEval Pack numbers (Measured across Code Generation, Explanation, and Debug tasks across the 6 most popular programming languages)

Source: IBM Research (May 2024)

InstructLab

contribute to and train
Large Language Models

Red Hat allows your business to **own your AI.**

Red Hat is enabling your business to run **your AI** in your way.

Red Hat's AI portfolio

Trust | Choice | Consistency

AI models	AI platforms	AI-enabled portfolio	AI workloads
RHEL AI	OpenShift AI	Red Hat Lightspeed	

Open Hybrid Cloud platforms

Partner ecosystem

Hardware | Accelerators | Models | Delivery

AI platforms

Red Hat Enterprise Linux AI

- Granite language and code models
- InstructLab
- open source AI tooling
- AI-optimized Linux

AI platforms

Red Hat OpenShift AI

Can be used for both GenAI and Predictive AI use cases

Helps accelerate getting AI apps and models into production

Provides a trusted platform, choice of AI models and hardware, and consistent operations

AI platforms

Red Hat Ansible Automation Platform

Automation is a key enabler for AI-driven workflows.

The content and structure of Ansible playbooks are a fertile and safe AI training ground.

Automation delivers the savings needed to support your AI transformation.

Red Hat Lightspeed

Neural Magic + Red Hat

Neural Magic

Neural Magic's expertise in software and algorithms that accelerate generative AI (gen AI) inference workloads.

vLLM

nm-vllm

LLM Compressor

DeepSparse

+

Red Hat

Red Hat's vision of high-performing, functional AI workloads that directly map to customer-specific use cases and data, anywhere and everywhere across the hybrid cloud.

Red Hat Enterprise Linux AI

InstructLab

Red Hat OpenShift AI

=

The potential to supercharge LLM deployments anywhere and everywhere across the hybrid cloud by providing a ready-made, highly-optimized, open inference stack.

Open source meets AI.
AI meets **Open Hybrid Cloud.**