



Vice-President of Engineering, Cohere

Rachad Alao

- Leading Agentic AI developments at Cohere
- Previously led AI Trust, Safety and Responsible AI at Meta Superintelligence Labs and Google Research.

Open Science and AI Foundation Models: Allies in the Pursuit of Knowledge

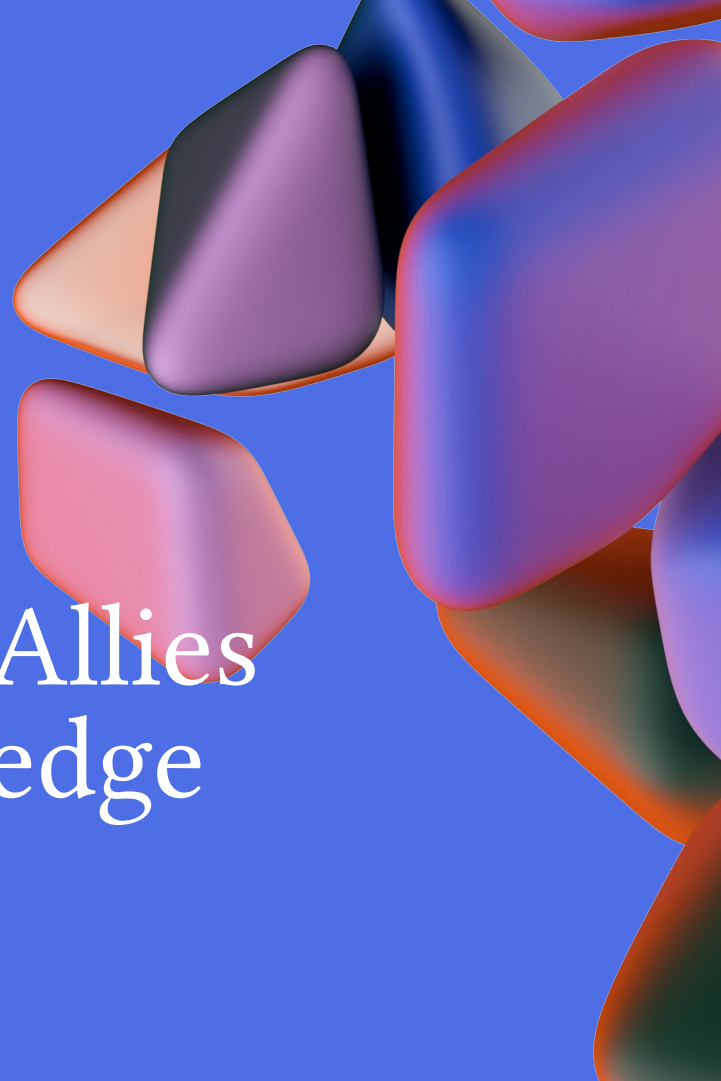


Table of contents

01 Speed of Innovation vs. Safety & Control

02 Transparency, Trust & Safety

03 Collaborative Innovation at PyTorch Scale

04 Democratization, Practical Scalability and Safety



Speed of Innovation vs. Safety & Control



Moving too fast and
breaking things?



**Anthropic says its AI
models hacked 3
organizations during
testing**



What should we do?

Stop everything?

Let a small group of Top AI Labs self-regulate?



Transparency, Trust & Safety



Transparency as the starting point

Transparency is essential to achieve Safety and Security

Models should not be black boxes

Evidence based Risk Framework



Testing, Auditing and Explainability

Auditability and testing by independent third parties

Open research enables broader auditability of models

Enables easier discovery of edge case vulnerabilities



Monitoring and Safeguards

Continuous monitoring of AI Systems

Dynamic Safeguards

Nvidia Open Agent Safety Platform



Collaborative Innovation at PyTorch Scale



Open Science, Open Source and Pytorch

Catalyst of AI revolution

Open source projects and open science artifacts are
key building blocks saving millions of GPU hours
globally



COHERE PRESENTATION DI

Pytorch
vLLM
Llama
Command A+
Tiny Aya
Transcribe
Inkling
Nemotron
torchcompile
torchitan
Safetensors

Accelerating Specialized Research

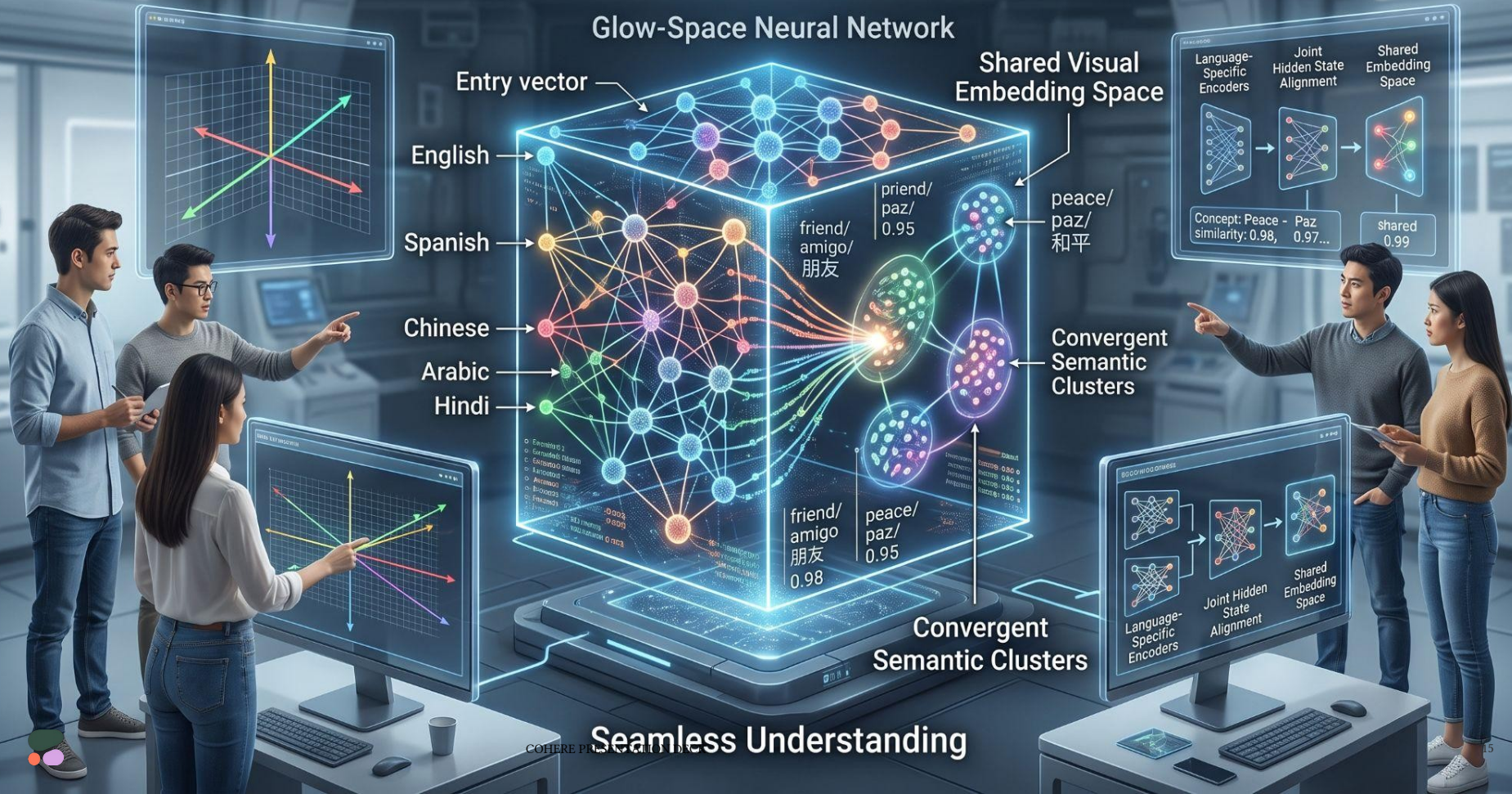
Access to internal representations enables researchers to push the frontier across many domains: Retrieval, Reasoning, Multilingual AI, etc,



Accelerating Retrieval Improvements



Cross-Lingual AI Alignment



Democratization, Practical Scalability and Sovereignty



Democratization, Model Accessibility

Open science is a crucial safeguard against centralized model access



A path to practical scalability

Agentic orchestration of Command, North mini code, Inkling, Nemotron, embed, rerank match frontier LLM agentic task success rate at a fraction of the compute



Enterprise Grade Sovereign, Safe and Secure AI

Cohere North, sovereign agentic AI stack for
enterprises built on open weights models





For more information,
please contact:

Rachad Alao

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

Thank you for listening!