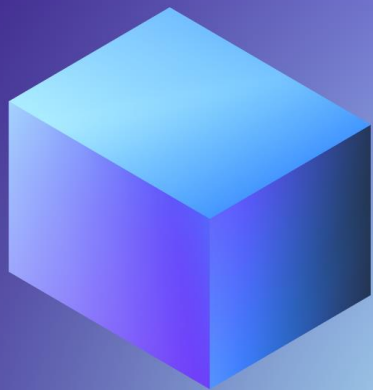aws

Powering Transformation:

How AWS and Pega Enable Businesses to Extract Real Value by Deploying Generative AI Responsibly

Daniel Aucoin, Sr. Partner Solutions Architect

aucoind@amazon.com

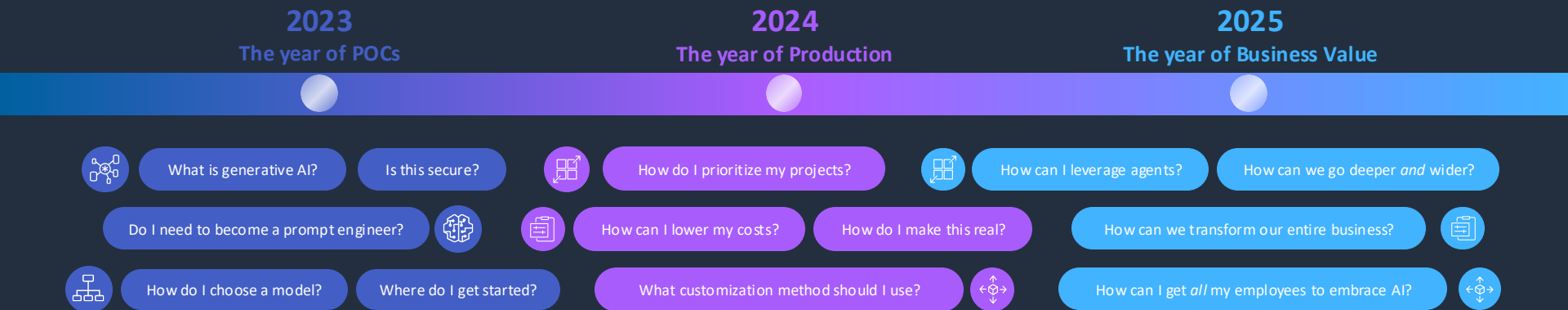Anubhav Sharma, Principal ISV Solutions Architect

shaanubh@amazon.com

Agenda

- Innovation opportunities

- Data is your differentiator

- How Pega uses AWS for Responsible, Secure Generative AI

- What's next?

Generative AI: Trends through the years

2023
The year of POCs

2024
The year of Production

2025
The year of Business Value

What is generative AI?

Is this secure?

How do I prioritize my projects?

How can I leverage agents?

How can we go deeper *and* wider?

Do I need to become a prompt engineer?

How can I lower my costs?

How do I make this real?

How can we transform our entire business?

How do I choose a model?

Where do I get started?

What customization method should I use?

How can I get *all* my employees to embrace AI?

Innovation opportunities

Generative AI has potential to create significant business value

NEW EXPERIENCES

Create new innovative and engaging ways of interacting with your customers and employees

PRODUCTIVITY

Radically improve productivity across all lines of business

INSIGHTS

Extract insights and clear answers from all your corporate information, enabling faster and better decisions

CREATIVITY

Create new content and ideas, including conversations, stories, images, videos, and music

Industry-specific Generative AI use cases

Healthcare & Life Sciences

Ambient digital scribe

Medical imaging

Drug discovery

Enhance clinical trials

Research reporting

Industrial & Manufacturing

Product design

Operational efficiency

Maintenance Assistants

Supply chain optimization

Equipment diagnostics

Financial Services

Portfolio management

Financial documentation

Intelligent advisory

Fraud detection

Compliance assistant

Retail

Pricing optimization

Virtual try-ons review

Marketing Optimization

Product descriptions

Pers. Recommendations

Media & Entertainment

HQ content at scale

Enrich broadcast content

Automated content tagging

Optimize subscriber experience

Automated highlights gen.

Nasdaq

Tackling insider dealing

NASDAQ leverages generative AI with Amazon Bedrock to enhance global market surveillance

CHALLENGE

Nasdaq needed to help clients more effectively monitor and detect potential market abuse. They wanted to enhance the quality, speed, and efficiency of market abuse investigations performed by clients.

SOLUTION

Nasdaq leverages Amazon Bedrock to create new kinds of intelligence reports for investors and corporate customers that leverage the company's proprietary data sets and drive faster, more impactful content creation in Nasdaq's marketing and communication teams. Nasdaq uses generative AI to expand its portfolio of financial Software as a Service (SaaS) products for identifying financial crime risks. Learn more.

OUTCOMES

○ Cuts investigation time by 33%

FINANCIAL SERVICES **ENTERPRISE** **NAMER** **USA** **AMAZON BEDROCK**

Data is your differentiator

Generative AI
Application

Generative AI Application

Data Foundation

WORKFLOWS

STORAGE

GOVERNANCE
& COMPLIANCE

DATABASES,
ANALYTICS,
& DATA LAKES

DATA INTEGRATION

Make generative AI work with **your data**

RETRIEVAL-AUGMENTED GENERATION (RAG)

Specialized knowledge through prompt augmentation

Enterprise knowledge corpus

No change to the foundation model

FINE-TUNING

Specialized knowledge for specific tasks

Small number of labeled examples

Change a copy of the foundation model

CONTINUED PRE-TRAINING

Generalized and specialized knowledge for your domain

Unlabeled, unstructured enterprise data

Change a copy of the foundation model

Security at the foundation

Security Considerations for Generative AI

COMPLIANCE & GOVERNANCE

The policies, procedures, and reporting needed to empower the business while minimizing risk

---

Create generative AI usage guidelines

Establish process for output validation

Develop monitoring & reporting processes

LEGAL & PRIVACY

The specific regulatory, legal, and privacy requirements for using or creating generative AI solutions.

Retain control of your data

Encrypt data in transit and at rest

Support regulatory standards

CONTROLS

The implementation of security controls that are used to mitigate risk.

Human-in-the-loop

Explainability & auditability

Testing strategy

Identity and access management

RISK MANAGEMENT

Identification of potential threats to generative AI solutions and recommended mitigations.

Threat modeling

Third-party risk assessments

Ownership of data, including prompts and responses

RESILIENCE

How to architect generative AI solutions to maintain availability and meet business SLAs.
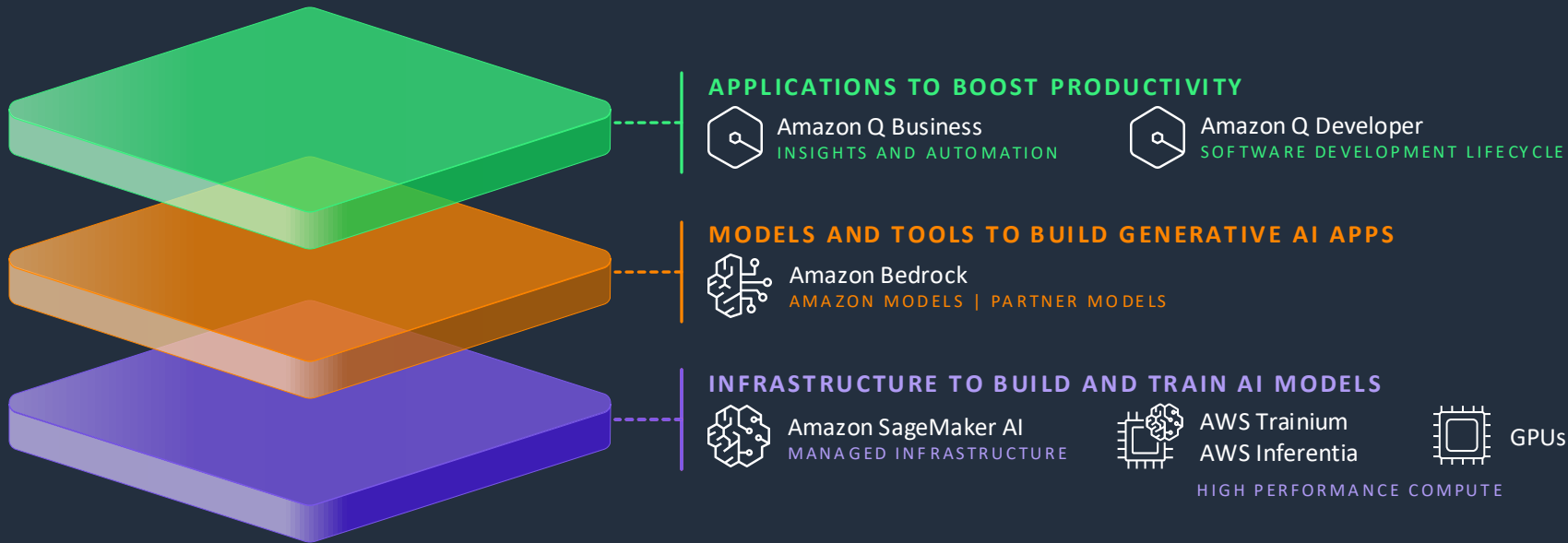

Data management strategy

Availability

High Availability and Disaster Recovery strategy

Amazon Bedrock

AWS Generative AI Stack

APPLICATIONS TO BOOST PRODUCTIVITY

Amazon Q Business
INSIGHTS AND AUTOMATION

Amazon Q Developer
SOFTWARE DEVELOPMENT LIFECYCLE

MODELS AND TOOLS TO BUILD GENERATIVE AI APPS

Amazon Bedrock
AMAZON MODELS | PARTNER MODELS

INFRASTRUCTURE TO BUILD AND TRAIN AI MODELS

Amazon SageMaker AI
MANAGED INFRASTRUCTURE

AWS Trainium
AWS Inferentia

GPUs

HIGH PERFORMANCE COMPUTE

Amazon
Bedrock

The easiest way to build and scale generative AI applications with foundation models (FMs)

Choice of leading FMs through a single API

Model customization

Retrieval Augmented Generation (RAG)

Agents that execute multistep tasks

Security, privacy, and safety

Amazon Bedrock keeps data secure & private

- None of the customer's data is used to train the underlying foundational models.

- All data is encrypted at rest using AWS KMS and encrypted in transit with TLS 1.2 (minimum). Also supports Amazon Managed keys or Customer-Managed Keys (CMKs).

- Fine-tuned models are encrypted and stored using customer AWS KMS key. Only you have access to your customized models.

- Data used to customize models remains within your VPC.

- Integration with IAM for fine-grained access controls, and GuardDuty for threat detection.

- Support for data privacy standards, including GDPR, HIPAA, and PCI

- SOC 1, 2, 3 Compliant

Amazon Bedrock

AI21labs	**amazon**	**ANTHROP\C**	**cohere**	**Luma**	**Meta**	**MISTRAL AI_**	**poolside**	**stability.ai**
Effective reasoning & rapid analysis for long context windows	Frontier multimodal intelligence at low-latency, Agent & RAG Applications, high-quality image & video generation	Advanced reasoning & coding capabilities, including computer use skills	Multimodal search & advanced retrieval powering multilingual knowledge agents	High-quality video generation from text & images	Advanced image & language reasoning	Knowledge summarization, expert agents, & code completion	Software engineering AI for large enterprises	High-quality AI image generation, easily deployable at scale
JAMBA	AMAZON NOVA	CLAUDE	COMMAND EMBED RERANK	LUMA RAY 2	LLAMA	MISTRAL MIXTRAL	MALIBU POINT	STABLE DIFFUSION STABLE IMAGE

Coming soon

Amazon Bedrock Marketplace enables developers to discover, test, and use over 100 popular, emerging, and specialized foundation models (FMs) alongside the current selection of industry-leading models in Amazon Bedrock.

Responsible AI builds trust

Responsible AI Dimensions

FAIRNESS

Considering impacts
on different groups
of stakeholders

EXPLAINABILITY

Understanding
and evaluating
system outputs

CONTROLLABILITY

Having mechanisms
to monitor and steer
AI system behavior

SAFETY

Preventing harmful system output
and misuse

PRIVACY & SECURITY

Appropriately obtaining,
using and protecting
data and models

GOVERNANCE

Incorporating best practices
into the AI supply chain,
including providers
and deployers

TRANSPARENCY

Enabling stakeholders to
make informed choices
about their engagement
with an AI system

**VERACITY &
ROBUSTNESS**

Achieving correct
system outputs,
even with unexpected
or adversarial inputs

European Parliament

CHALLENGE

The European Parliament wanted to make millions of official documents in paper archives accessible to researchers, policymakers, educators, and the public.

SOLUTION

They built Archibot, a sophisticated AI assistant to offer users worldwide easy access to EU Parliament documents in multiple languages - including resolutions, positions, policies, and inter-institutional negotiations - while also helping researchers and staff analyze data and create comprehensive reports. The project showcases how generative AI can be used by governments to enhance transparency and accessibility for citizens. What once required deep searches into the EU's physical archives is now available instantly, bringing the story of European democracy to life for citizens everywhere. Learn more.

OUTCOMES

o   Made over 2 million documents accessible

BRINGING HISTORY TO LIFE

European Parliament used Anthropic's Claude on Amazon Bedrock to use AI in a controlled and responsible way and launch Archibot, making over 2 million documents accessible

---

GOVERNMENT   ENTERPRISE   EMEA   FRANCE   AMAZON BEDROCK

Responsible AI:
Best practices

Put your people first

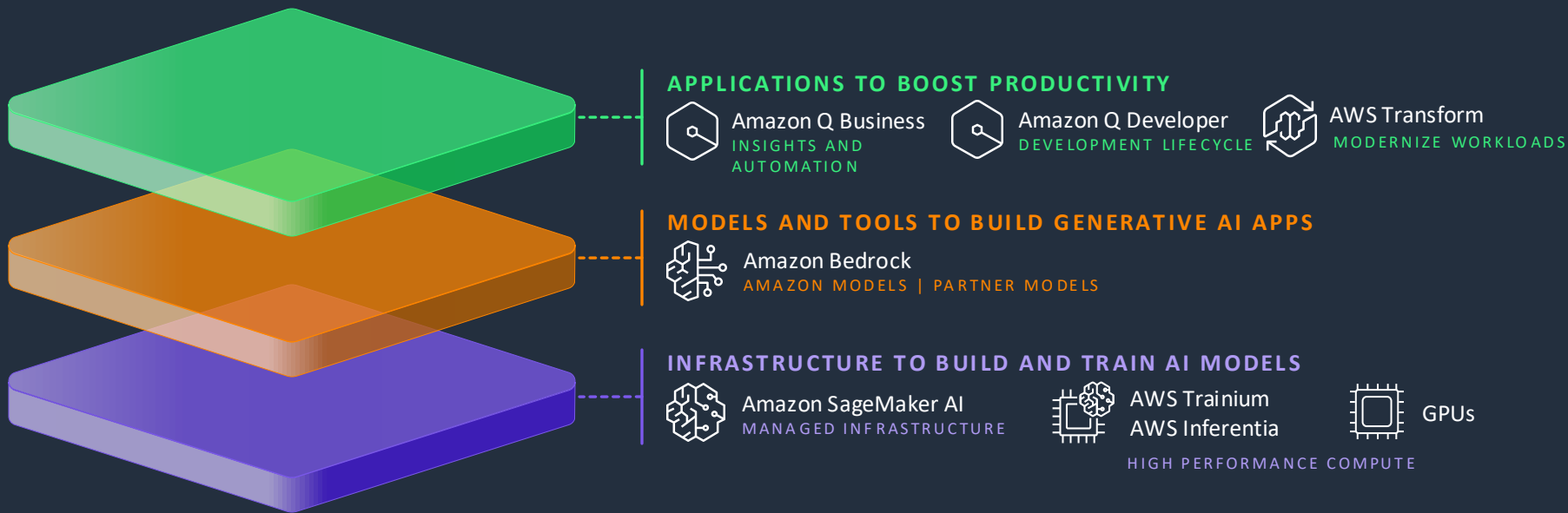Assess risk on a (use) case-by-case basis

Iterate across the AI lifecycle

Test, test again, and then test again

AWS Transform for Mainframe

AWS Generative AI Stack

APPLICATIONS TO BOOST PRODUCTIVITY

Amazon Q Business
INSIGHTS AND AUTOMATION

Amazon Q Developer
DEVELOPMENT LIFECYCLE

AWS Transform
MODERNIZE WORKLOADS

MODELS AND TOOLS TO BUILD GENERATIVE AI APPS

Amazon Bedrock
AMAZON MODELS | PARTNER MODELS

INFRASTRUCTURE TO BUILD AND TRAIN AI MODELS

Amazon SageMaker AI
MANAGED INFRASTRUCTURE

AWS Trainium
AWS Inferentia

GPUs

HIGH PERFORMANCE COMPUTE

State of enterprise workload transformation

70%

of workloads
still on premises
Gartner

70%

of legacy IT software for
Fortune 500 companies
written 20+ years ago
McKinsey

1.5 years

average time to
complete a migration ISG
Research

Enterprises face many transformation challenges...

Slow transformation

Complex legacy systems

Labor-intensive processes

Limited resources

Limited scalability
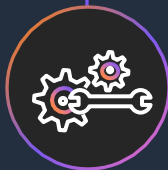
Sequential processes

Limited collaboration

Monolith systems that resist modular change

Scarcity of expertise

Shrinking pool of legacy system experts

Specialized hybrid skills requirement

AWS Transform
agent for mainframe

The first agentic AI service for modernizing mainframe workloads

Automate the entire transformation process—from initial analysis and planning to code refactoring and migration

Reduce mainframe modernization timelines from years to months

Analyze millions of lines of COBOL

Generate comprehensive documentation, and code analysis in minutes

Toyota Motors North America

TOYOTA

CHALLENGES

- Key business logic trapped in thousands of legacy programs
- Shrinking mainframe skills pool
- Direct Model usage did not yield better results
- 18,000 MIPS
- 40M+ lines of COBOL code

SOLUTIONS

- Leverage AWS Transform to perform Analyze tasks with agents, identify missing components and prepare a stable code base for modernization
- Create domains and plant seeds for decomposition
- Generate XML and PDF detailed documentation
- Leverage Amazon Q Business for Chat assistance on Documentation

RESULTS

More than 50% faster Mainframe modernization

Application discovery and planning accelerated by 75%

Reduced application domain knowledge extraction efforts from months to hours and 90% acceleration

Decomposing monolithic application into business domains with simple user inputs
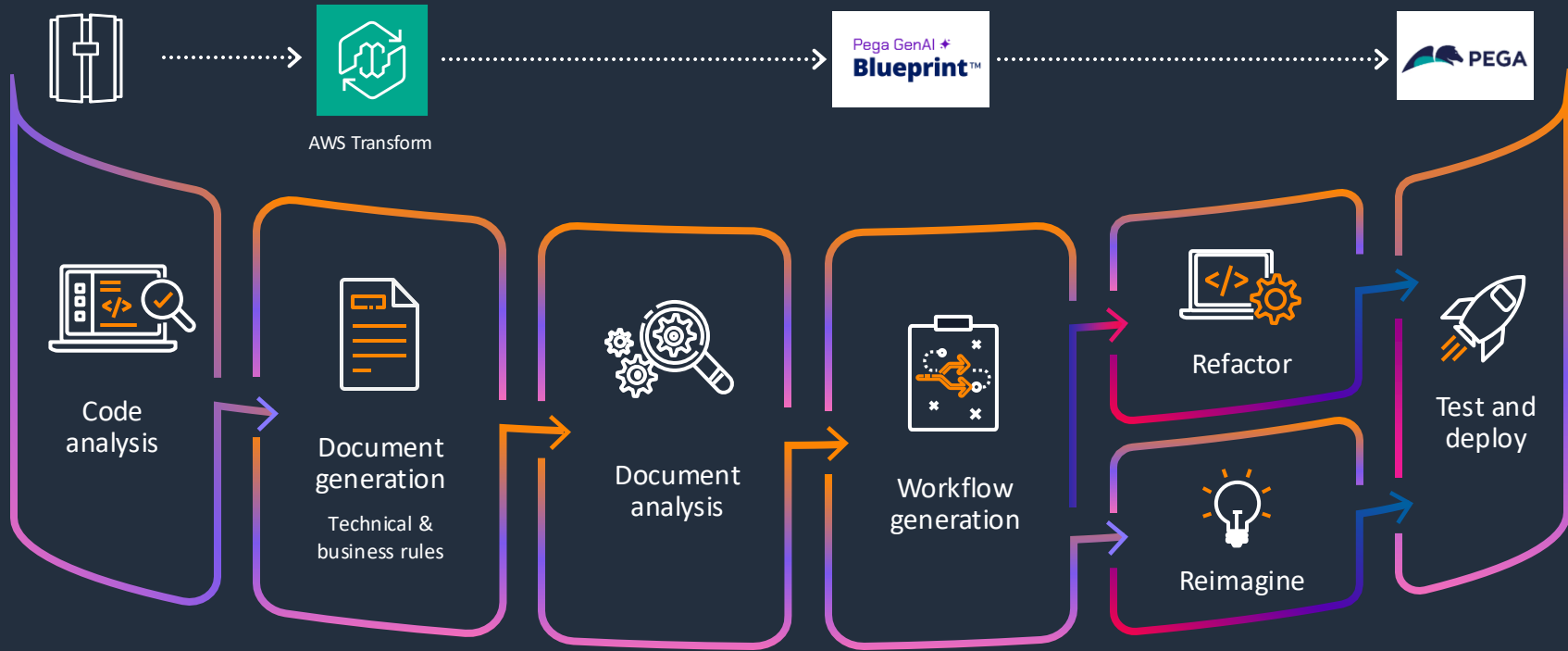
INDUSTRY
Automotive

REGION
Global

Toyota (NYSE:TM) has been a part of the cultural fabric in the U.S. for nearly 70 years, and is committed to advancing sustainable, next-generation mobility through our Toyota and Lexus brands, plus our nearly 1,500 dealerships. Toyota directly employs nearly 48,000 people in the U.S. who have contributed to the design, engineering, and assembly of more than 35 million cars and trucks at our 11 manufacturing plants.
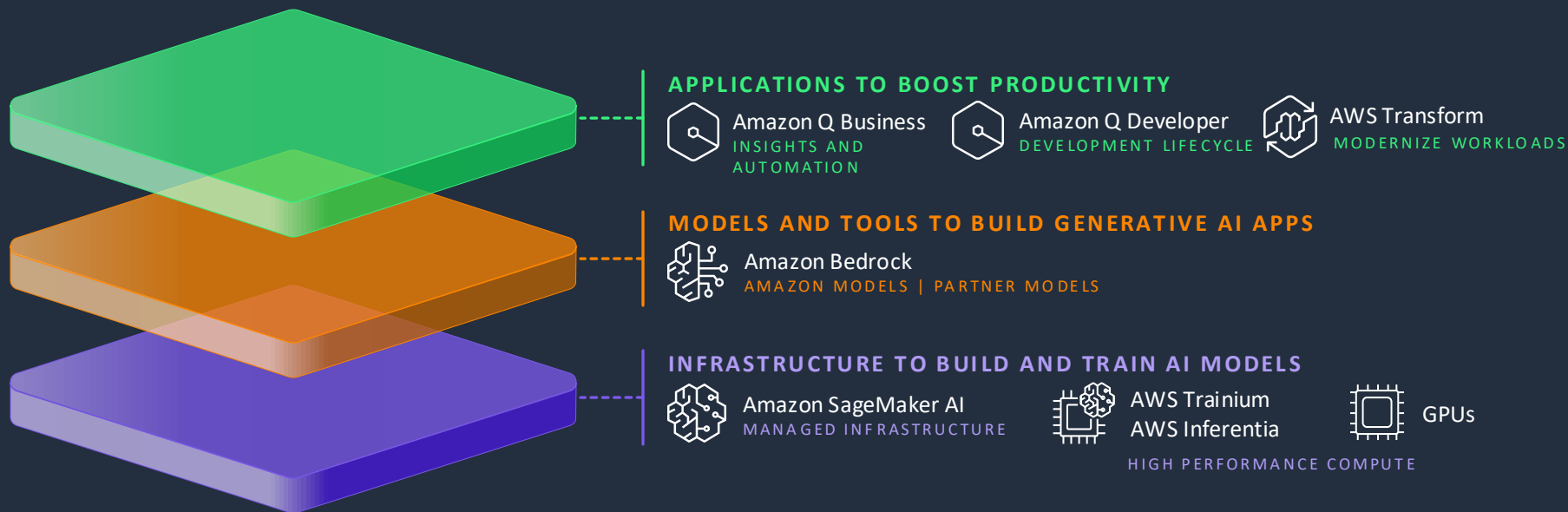
"AWS Transform has been a revolutionary solution for us. The powerful documentation, code analysis, and decomposition capabilities provided by AWS give us a clearer understanding of our application landscape. We're able to generate detailed documentation of hundreds of COBOL modules, including context-aware information and persona-driven insights, in less than a day. AWS Transform has done what many said was impossible.

Brian Kursar, CTO Toyota Motor North America

Mainframe Modernization with AWS Transform and Pega

Q Business Index

AWS Generative AI Stack

APPLICATIONS TO BOOST PRODUCTIVITY

Amazon Q Business
INSIGHTS AND AUTOMATION

Amazon Q Developer
DEVELOPMENT LIFECYCLE

AWS Transform
MODERNIZE WORKLOADS

MODELS AND TOOLS TO BUILD GENERATIVE AI APPS

Amazon Bedrock
AMAZON MODELS | PARTNER MODELS

INFRASTRUCTURE TO BUILD AND TRAIN AI MODELS

Amazon SageMaker AI
MANAGED INFRASTRUCTURE

AWS Trainium
AWS Inferentia

GPUs

HIGH PERFORMANCE COMPUTE

The current state of "future work"

**The average desk worker uses 11 applications to complete their tasks,
up from just 6 in 2019.**

- **Gartner**
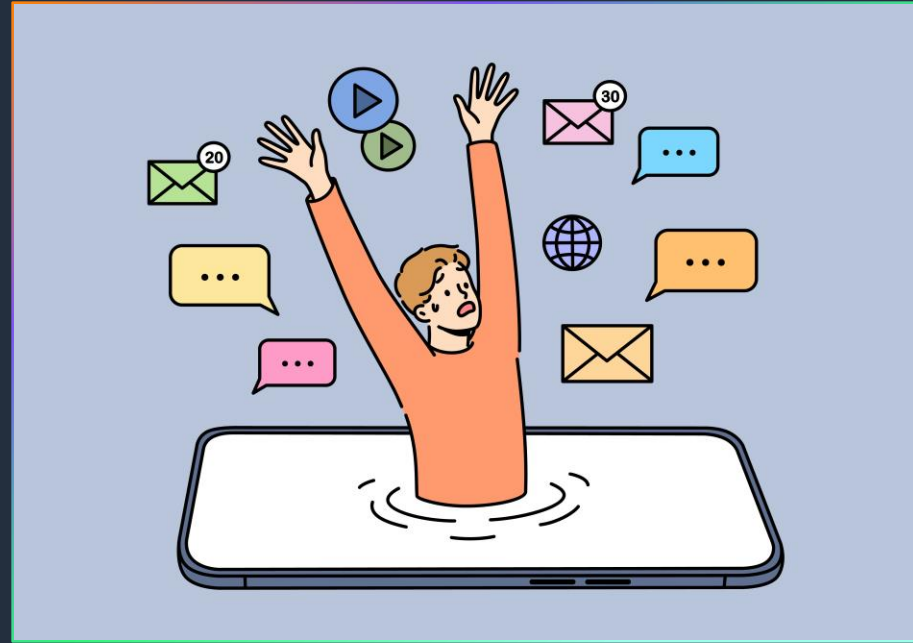
Reimagining productivity

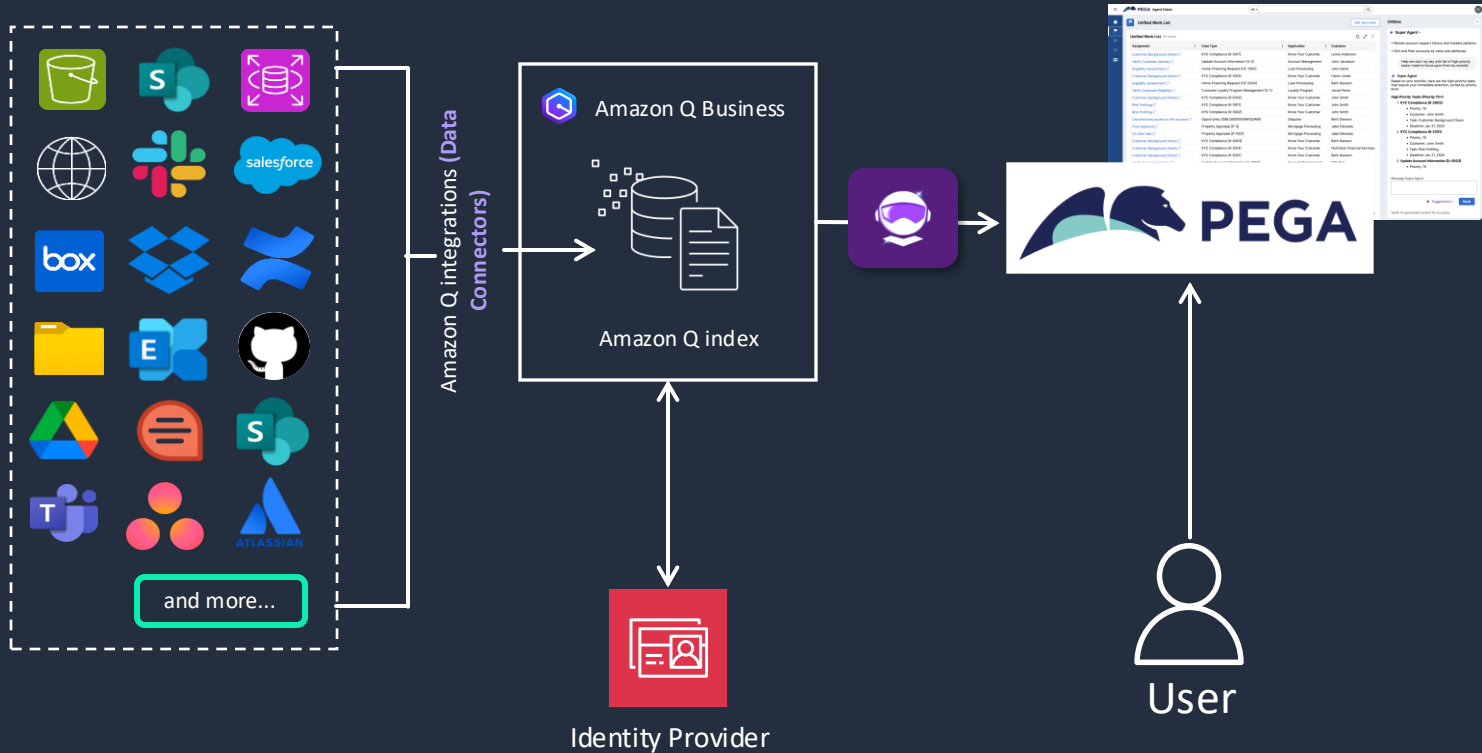Utilize cross-app insights with Amazon Q Business Index

Amazon Q index
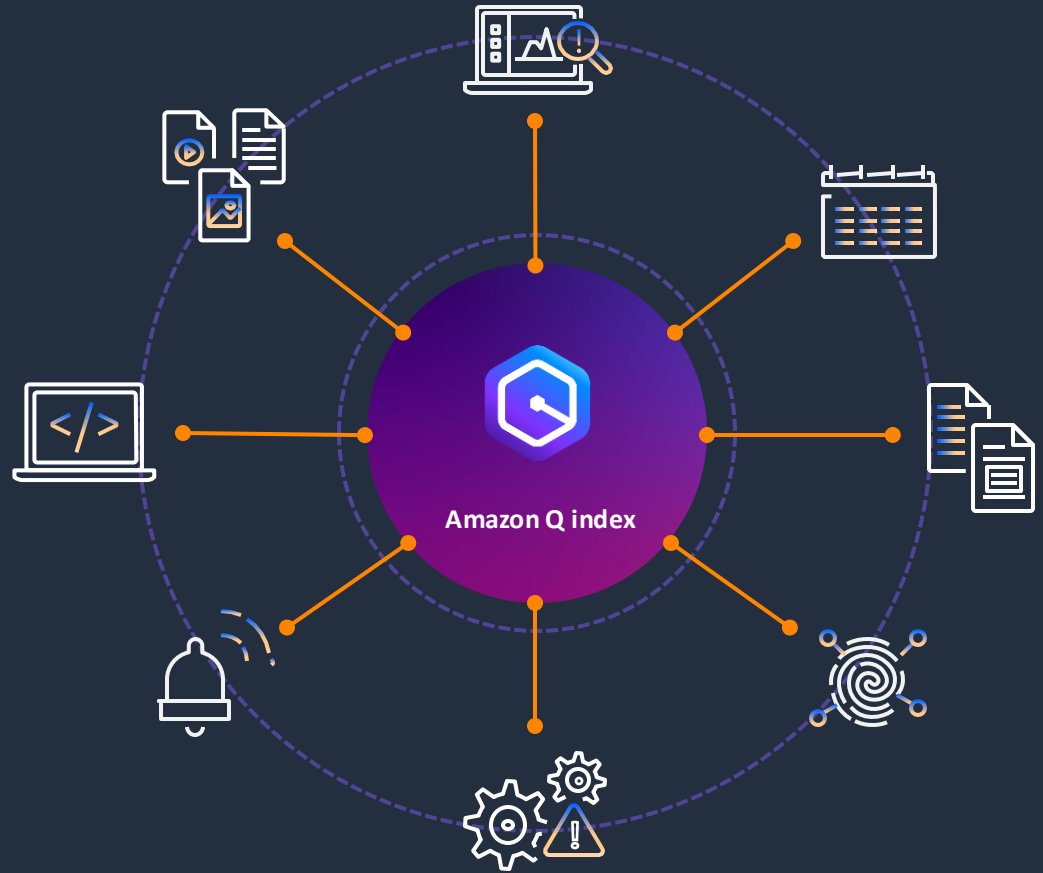
Amazon Q Business customers

Enriched GenAI Experiences Using Q Business Index

What does **the future** hold for generative AI?

Agents

Multi-modal

Multiple models

AI Policies & Standards

Strategic recommendations on generative AI

EYES WIDE OPEN

- Culture of continuous experimentation

- Prevent early dependencies

NO ONE SIZE FITS ALL

- Generative AI as an enhancement – not a replacement.

- Requirements in enterprises differ based on financial resources, security needs, governance and skills.

- Evaluate buy-vs-build

FLEXIBILITY IS KEY

- Innovation requires flexibility free from technical or contractual lock-ins.

- Infrastructure supporting 3rd-party Generative AI integration

- A breadth of services ensures long-term flexibility and business value

LISTEN AND ENABLE

- Work backward from your customers

- Listen to your domain experts

- Enable employees with the right set of tooling

aws
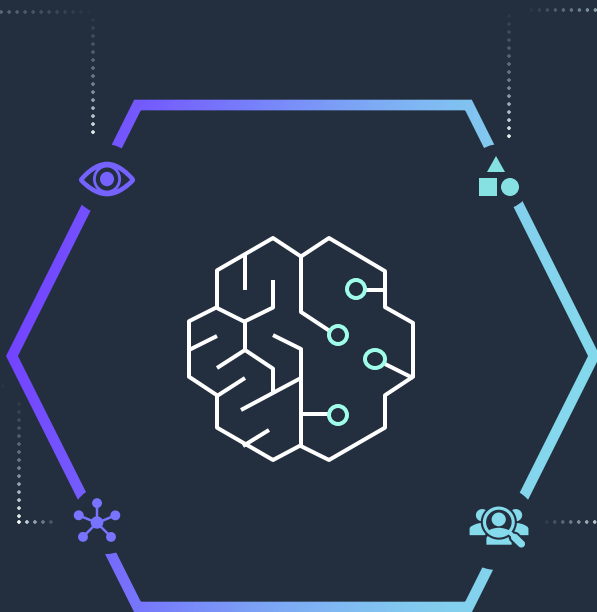
Thank you!

Daniel Aucoin, Sr. Partner Solutions Architect

aucoind@amazon.com

Anubhav, Principal ISV Solutions Architect

shaanubh@amazon.com