IBM Consulting

The generative AI opportunity for the automotive industry

Table of contents

01

Thriving in a new technological environment

02

How generative AI can help the automotive industry

03

The opportunity for automotive companies

04

Scaling generative AI in the automotive industry

05

The IBM approach to generative AI-based transformation

06

Implications for the automotive industry and how IBM can help

01
Thriving in a new technological environment

The world of AI has witnessed ground-breaking advancements in recent years, and generative AI has rapidly become one of the most promising developments in the decades-long evolution of AI technologies. Automotive companies globally are mobilizing to seize opportunities and respond to emerging challenges associated with this new technology. Generative AI is evolving rapidly and becoming increasingly accessible through applications such as ChatGPT and DALL·E 2. Along with increased focus and interest, we are also seeing increased investment activity as automotive leaders continue to evaluate and experiment with this new capability—to assess both opportunities and risks.

At the World Economic Forum 2024 in Davos, IBM Chairman and CEO Arvind Krishna projected an annual productivity gain of about USD 4 trillion from the use of AI by the end of the decade.[1] Such growth could prove to be a staggering boost to the economic competitiveness for companies and nations, and those that embrace AI would be at an advantage. As a result, companies are building infrastructure, data, workforce and business networks to harness these benefits. Of particular importance is the need for a workforce with sharpened critical thinking skills, which will better position individuals to use AI to improve their own capabilities while also enhancing AI outputs. As AI continues to shape all sectors, the emphasis on these skills will reinforce the need for an adaptable workforce equipped to work with increasingly sophisticated technologies.

The automotive industry is at a historic turning point. Rapid electrification of the industry is coinciding with the move toward software-defined vehicles. These tectonic shifts are fundamentally changing the dynamics of the industry with new players emerging and old ones vying for improved positions. With the industry in flux, we believe generative AI can be a key factor in determining which companies will succeed in the future makeup of the automotive industry.

In addition to these changes in the vehicles, we're also witnessing a shift in how cars are sold and how automotive companies connect with their end customers. The traditional dealer model is being replaced with automotive companies engaging directly with consumers by selling vehicles to them or providing access to vehicles through various leasing or subscription-based models. Generative AI holds great potential for helping automotive companies effectively personalize their interactions with individual customers.

Technology is at the core of business transformation for automotive companies and generative AI has the potential to materially improve the productivity and speed at which technology is built and deployed enterprise wide. Automotive companies' existing investments in digital, data and cloud technology provide a strong foundation to harness the potential of generative AI.

In the next section, we discuss what automotive companies are doing with generative AI and the types of use cases that show the greatest promise in the automotive industry.

02

How generative AI can help the automotive industry

IBM is currently working with automotive original equipment manufacturers (OEMs) across the globe to apply AI solutions to help execute transformational programs. Early adopters of generative AI technologies are already realizing benefits through operational efficiencies, productivity gains, speed of action and risk avoidance—with significant operational cost savings in selected areas.

We believe generative AI can help automotive companies enhance productivity across all functional areas, from product development to marketing to after-sales service.

In product development, for example, we observed automotive companies starting to apply generative AI throughout their entire software development lifecycle, from user story creation to coding to testing. Implementing AI solutions has significantly reduced their development cost and time-to-market and improved quality and standards compliance.

In vehicle production, quality and uptime are critical measures for automotive OEMs. IBM Research® is doing groundbreaking research in generative AI foundation models for time series data. The research shows promising results in predicting quality issues and equipment failures which can enable OEMs to address many such failures before they happen.

In marketing, generative AI can improve productivity significantly, as marketing copy that is compliant with brand guidelines can now be automatically generated. Generative AI also provides an unprecedented possibility to truly personalize marketing messages for individuals.

In after-sales service, the rapid shift to electric and software-defined vehicles is challenging the skills of service technicians. Generative AI can provide natural language instructions and step-by-step guidance to these service technicians, which can help bridge the skills gap and improve productivity.

Given early successes in several areas, leading automotive OEMs are investing heavily to institutionalize generative AI at scale. As seen in Figure 1, the areas that are currently seeing the most use of generative AI include product development, customer engagement, operational efficiency and technology modernization.

Figure 1. Areas that are seeing the most use of generative AI

Customer engagement

Virtual agent
Enable virtual agents to generate context-aware and realistic responses to customer queries.

Conversational AI
Find answers with natural responses generated from company knowledge bases.

Operational efficiency

Personalized marketing
Personalize marketing and campaign collateral and pricing to drive increased conversion.

Technician assistant
Deploy a conversational AI chatbot that advises service technicians on finding the right solution and providing step-by-step instructions using company-specific data.

Product development

Software development
Speed up the entire software development lifecycle from requirements definition through code generation to testing and documentation.

Achieve compliance with industry standards by design.

IP protection
Analyze new ideas against existing patent data to identify patentable ideas as well as potential infringement risks.

Generative design
Speed up product design by generating alternative design ideas that the human designer can choose from and develop further.

Technology modernization

Code modernization
Modernize existing code while simplifying system context, design and redundancies.

IT development
Drive the entire lifecycle management of IT development from ideation to code, including user story development, design, code and test case generation.

Code generation
Generate cloud-native code using natural language for Java, Quarkus, Springboot, NodeJS, Python, Red Hat® Ansible® and many others.

Generative AI can target work-intensive tasks, analyze large datasets, interpret complex documentation and synthesize these findings into actions. Many automotive companies are actively attempting to quickly learn and adopt generative AI concepts. However, only a few OEMs are using a structured approach to embrace generative AI, which is enabling them to scale solutions rapidly into production and gain an enterprise-wide advantage.

IBM employs an automotive industry-specific operating framework called the component business model (CBM). IBM has utilized this model with automotive OEMs worldwide to identify priority areas for action. The model can be rapidly applied to any organization's people, processes and technology capabilities. The goal is to reveal areas of resource and cost concentration, misalignment of strategic directions in the operating model, redundancies, duplication and gaps in execution capabilities.

This deterministic approach takes into consideration the current maturity of generative AI capabilities, the readiness of generative AI toolsets, risk exposures, data sharing and security, and the ability of enterprises to take any proof of concept to production deployment. Figure 2 combines all these factors and depicts our current view of an automotive company's priority areas for the application of generative AI to derive business advantages.

The IBM CBM framework enables faster outcomes by helping create focus and consensus rooted in organization-specific operational facts and technology considerations.

Figure 2. The CBM framework for automotive OEMs

	Product and process development	Supply chain	Production	Marketing and sales	Services and aftersales	Digital services	Financial management	Business administration
Strategy and direct	Research and development ●●	Demand planning ●	Master production planning ●	Brand management	Post sale strategy	Service development and monetization strategy	Capital appropriation planning	Organization and process policies
	Design rules and policies ●	Supplier relationship planning	Production strategy	Sales and promotion planning ●	Software update strategy ●	Service provider strategy	Financial planning and forecasting	Alliance strategies
	Portfolio strategy and planning ●	Supply chain strategy and planning	Production rules and policies	Consumer relationship strategy		Platform strategy		Corporate and line of business strategy
						Data strategy		
Manage and monitor	Design validation ●●	Supply chain performance monitoring	Production scheduling and monitoring ●	Consumer relationship monitoring ●	Quality management	Service development management	Treasury	Legal and regulatory
	Configuration management ●●	Supplier management ●	Quality control ●	Dealer management ●	Warranty management	Platform management	Tax management	Business performance
	Change management ●	Logistics management ●		Demand forecast and analysis	Software update management ●●	Service delivery management ●	Risk management and internal audit	Human capital management
	Program management ●					Data security management ●		Intellectual property ●
Execute and deliver	Tool design and build ●●	Inventory management ●	Plant operations ●	Consumer interaction management ●	Parts management	Service development	Accounting and general ledger	Indirect procurement
	In-vehicle system design ●●	Direct procurement ●	Maintenance management ●	Financing and lease management	End-of-life management	Service localization ●	Cost management	IT systems and operations
	Process design ●	Transportation management ●		Order management	Vehicle service management ●	Support center operation ●●		Facilities and equipment management
	Mechanical and electrical design ●				Software update delivery ●●	Data maintenance ●		Knowledge and learning ●
	Software development ●							

Generative AI use cases
- Customer engagement
- Operational efficiencies
- Product development
- Modernization

Impact of generative AI
- High: > 30% expected productivity impact
- Medium: 10%–30% expected productivity impact
- Low: < 30% expected productivity impact

03
The opportunity for automotive companies

Generative AI has the potential to revolutionize the way automotive companies operate, and we're just beginning to unearth the wealth of opportunities. A recent study conducted by the IBM Institute for Business Value (IBV) identified three key categories where automotive companies can realize benefits from adopting generative AI: expanded capabilities, business growth and cost reduction (see Figure 3).

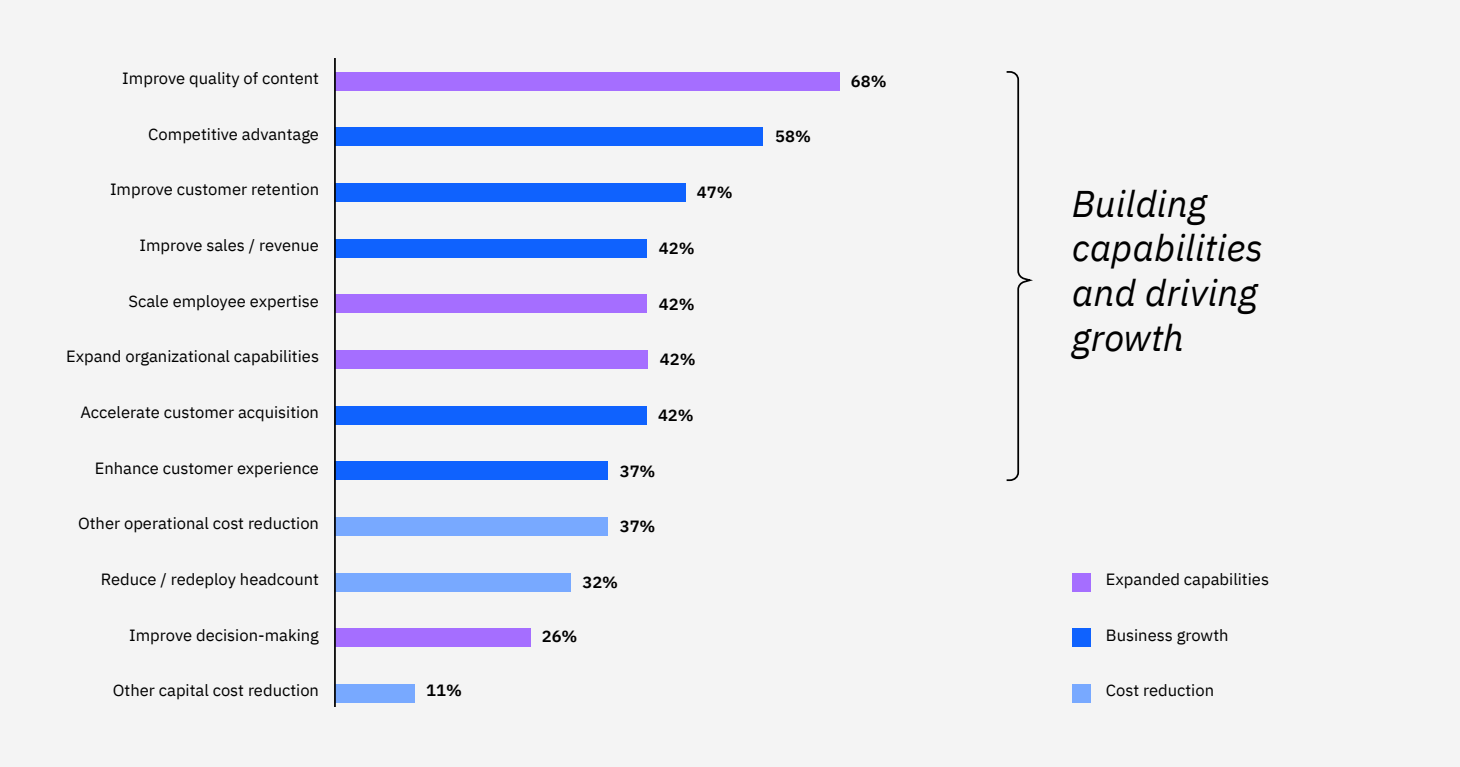

Figure 3. Benefits of adopting generative AI in automotive[2]

We have identified a few use cases that could quickly transform your AI efforts and deliver a substantial and sustainable ROI. This list is not intended to be exhaustive—it's based on our understanding of a typical automotive company's operations, our observations of successful market trends and what we believe is ready for implementation. We recognize that some of these use cases might be emerging and experimental.

We expect this list to evolve and expand as automotive companies embark on this journey. We will discuss this long and fruitful journey further in the following section.

Generative AI use cases for the automotive industry

Generative AI use cases
- Customer engagement
- Operational efficiencies
- Risk management
- Technology modernization

Priority areas for generative AI

Customer engagement	Generative AI application	Potential impact	Where to start
Customer interaction management	Drive cost and time efficiencies into the content supply chain using large language models (LLMs) for content generation and asset derivative production. Improve search engine optimization and campaign channel deployment using LLMs to analyze web pages and suggest improvements to enhance organic SEO. Dynamically create personas and journey maps for highly tailored segmentation.	Achieve significant reductions in content creation spend. Improve engagement rates with social media assets.	Deploy generative AI-enabled copy production for marketing campaigns. Deploy enterprise-safe image content generation technology while training an internal LLM on brand and campaign guidelines, personas, brand voice and more.
Dealer management	Automate the auditing of dealer websites using LLMs to ensure the accuracy of information and compliance with your brand guidelines.	Provide a strong risk management solution for an automotive company's brand representation and reduce associated legal risks. Audit all dealer websites instead of sampling them, reduce manual labor by auditors and provide corrective recommendations to dealers.	Deploy generative AI to audit dealer websites against company-specific rules loaded into the LLM. Create an automated workflow that focuses human auditors' time on sites flagged as noncompliant by generative AI.

Product development	Generative AI application	Potential impact	Where to start
Software development	Speed the entire software development lifecycle from requirements definition through code generation to testing and documentation. Achieve compliance with regulatory requirements and industry standards by design. Generate synthetic data to train and test autonomous driving and advanced driver assistance models and applications for various scenarios.	Improve software developer productivity. Improve software quality and reduce time to quality. Improve compliance with regulations and standards and reduce the cost of compliance. Reduce the cost of collecting training and test data, particularly for less common weather conditions.	Develop “software engineering agents” to help speed specific tasks along the software development lifecycle. Incorporate company-specific software engineering knowledge to develop differentiating agents representing your policies and values.
Intellectual property	Analyze new ideas against existing patent data to identify patentable innovations and potential infringement risks. Find relevant passages in existing patents and give recommendations that result in successful patent applications.	Improve the ability to productize and monetize innovation. Reduce the risk of IP-related issues and the cost of patent litigation.	Train your LLM on a selected set of existing patents. Create a workflow to automatically compare invention disclosures against existing patents.
Mechanical and electrical design	Speed product design by generating alternative design ideas that the human designer can choose from and develop further. Insert engineering and regulatory constraints into the generative design process to reduce the number of iterations required to reconcile design and engineering constraints.	Enable faster time-to-market for new products. Generative AI can assist with work from concept to detailed design, so human designers spend less time creating concepts and refining the details. Reduce the number of iterations to incorporate engineering and regulatory constraints, such as aerodynamics and safety, into the design.	Develop a library of prompts to help designers achieve the desired outputs. Incorporate engineering and regulatory constraints into the model to speed iterations. Incorporate the company’s design language—to the extent that it’s codified—to improve the quality of output.

Operational efficiency	Generative AI application	Potential impact	Where to start
• Vehicle service management • Warranty management	Use a conversational AI chatbot that advises service technicians on finding the right solution and providing step-by-step instructions using company-specific data.	Improve the quality and consistency of customer experiences at the dealership. Reduce warranty cost by enabling service technicians to focus on the right issues.	Apply generative AI as a knowledge search engine, enabling service technicians to operate based on the best available information. Expand generative AI to act as a virtual assistant at the dealership, providing consistent guidance on next actions.
• Human resources management	Enable end-to-end inquiry digitization using generative AI for classification, knowledge search and highly accurate request integration to enhance HR capacity and facilitate employee empowerment. Use generative AI across the recruitment lifecycle to enable extreme automation of talent identification and onboarding, which can result in seamless and efficient outcomes.	Increase first point of contact (FPOC) query resolution.	Apply generative AI as a virtual assistant across the HR servicing domain, delivering proven value at low risk. Expand to knowledge search for policy and guidelines alongside increased integration to platforms such as ERP and ITSM systems. Address the recruitment lifecycle with onboarding campaigns augmented by generative AI and automation.
• Plant operations • Quality management	Use generative AI algorithms based on historical time series data to predict quality issues and unplanned maintenance events.	Improve manufacturing line uptime through better predictive maintenance. Reduce rework by identifying emerging quality issues early on.	Use historical time series data to train a foundation model to predict quality and maintenance events.

Operational efficiency	Generative AI application	Potential impact	Where to start
IT systems and operations	Technology and application modernization is an area ripe for transformation with generative AI. Examples of modernization areas include: **Building software from a collection of prompts** Drastically reduce the time and cost of building applications to enable your organization to move faster, with reduced costs and increased agility. **Converting software code from one language to another** Modernize aging systems developed in older generation languages to be more modular with modern languages or microservices. Take advantage of the latest development tools and processing capabilities for richer and more efficient software outcomes. **Creating automated test scripts** Automate the generation of test scripts for newly created or augmented application code for increased agility.	Facilitate faster software development lifecycle and accelerate new products and feature functionality releases. *Note: IBM is seeing that ~60% of our Ansible playbooks for applications can be generated based on natural language prompts. Additionally, we are observing 20%-30% productivity gains in the creation and maintenance of these playbooks.*	Deploy generative AI tools such as IBM® watsonx Code Assistant™ and automated test scripts to accelerate modularization and modernization of the core application portfolio. Generate playbooks and provide valuable problem-resolution recommendations based on data analysis and pattern recognition.

While individual use cases will continue to emerge as organizational confidence and generative AI technologies develop, another critical component of success will be the operating model needed to deploy these technologies consistently and safely across the business. In the following section, we discuss what automotive companies will need to do to scale and sustain investments in generative AI and how IBM can support them.

Case studies

Automatic creation of marketing copy
Marketing is one of the areas that stands to undergo the largest transformative changes because of generative AI, but also benefit the most from it.

> In a study by the IBV, 76% of chief marketing officers (CMOs) believe generative AI will change the way marketing operates. They also believe the failure to quickly adopt generative AI will significantly hurt their ability to stay competitive.[3]

We worked with the UK national sales company of a global automotive OEM to transform and automate its marketing content supply chain. Previously, creating marketing copy was a manual task. Reviewing the resulting content against the OEM's tone-of-voice guidelines was also manual and time-consuming.

Generative AI helped automate this process almost completely. The team trained a generic LLM to act as a marketer and ingested the OEM's tone-of-voice guidelines. The resulting assistant is now able to take the technical features of a vehicle model as input and use them to produce compelling marketing copy that is compliant with the guidelines by design. Such automation speeds the creative process tremendously, eliminates the need for a manual review of the output and handles factual and grammatical errors and tone issues that might have previously slipped through the manual review.

The client's CMO commented the following on the quality of the results: "I can take two versions of the same marketing copy, and I can usually immediately tell which one is created by generative AI. It's the one that doesn't have errors in it."

Prioritization of patent applications
Maintaining a strong patent portfolio is critical for automotive companies to protect their unique competitive advantage. Reviewing innovation disclosures and deciding what can and should be patented, however, can be resource-intensive and time-consuming. We worked with a global automotive OEM, that had a backlog of 4,000 innovation disclosures that needed to be prioritized and analyzed. Only then could we determine whether the prior art exists for a proposed innovation or if there's potential to patent it.

The IBM team built a generative AI solution that does most of the initial prioritization and screening. We trained a LLM on a large collection of patents in the areas of interest, and the key features of those patents were automatically extracted. Next, the innovation disclosure document was passed through the same feature extraction engine and the features of the innovation disclosure were compared to the features of existing patents. If similar features appeared to be covered by an existing patent, then the model advised the patent researcher of the relevant sections in existing patents. The model also highlighted why the existing patent appeared to cover the disclosed innovation.

The human patent researcher still makes the final decision as to whether an innovation is novel and should be patented or if prior art exists and the innovation is not patentable. However, the generative AI model speeds the work of a patent researcher quite substantially and allows them to spend more of their time with innovators. For example, they can advise the innovators on patent language.

Here's a comment from the patent research team lead: "You know, where I come from, we don't normally say anything when something is good. Because that's enough positive feedback. We only complain when something is not good. But what you have done in this short amount of time is really amazing and awesome! And after seeing your presentation and demo, I have so many new ideas about what we can do."

04
Scaling generative AI in the automotive industry

The "2024 CEO Study: 6 hard truths CEOs must face" from the IBM Institute for Business Value highlights 4 core elements that make up the generative AI opportunity canvas. These elements include increasing productivity across the enterprise; delivering more personalized, frictionless customer experiences; driving new revenue streams and business models; and transforming technology development and delivery.

When scaling generative AI, the question of how to safely harness it is as important as where to apply it. There are several scaling considerations for building trusted AI, including data privacy, IP, transparency and explicability, risks from early lock-in to any given model, compute and carbon costs, skills scarcity and, most importantly, governance. In the automotive industry, generative AI use cases will be varied and wide. However, to scale the deployment of generative AI technology safely and rapidly across different use cases, you need an operating model that delivers the necessary governance, control and skills. For this reason, establishing an AI Center of Excellence (CoE) is an essential first step. In our view, the AI CoE should establish the guardrails for trusted AI while accelerating innovation. Its key components are shown in Figure 4.

Here are some questions an automotive company should consider.

- What generative AI capabilities already exist in the organization to explore and test?
- What is the organization's appetite for developing talent and capabilities internally rather than engaging external partners and suppliers?
- How well does the organization's innovation culture, executive leadership, mission and risk tolerance align with its desire to innovate?
- What are the executives' expectations of ROI from innovation efforts? What is the approximate investment of time and money? What are their expectations around KPIs, milestones and time horizon?
- How receptive is the organization to open-source technologies and collaboration? How familiar is the organization with open-source tools?
- Are the criteria and KPIs understood regarding how to build trusted AI, such as data privacy, IP, transparency and explainability, risks from early lock-in to any given model, compute and carbon cost, skills scarcity and, most important of all, governance?
- Have the use cases been identified? Can the desired end state and value about how AI technology drives specific outcomes be described clearly and succinctly?
- If the use cases are not yet known, which innovation processes will be used to identify them? What are the criteria for selecting the use cases?
- Which person or role in the company is ultimately responsible for the operating model, governance and controls for safely and rapidly scaling the generative AI technology across use cases?

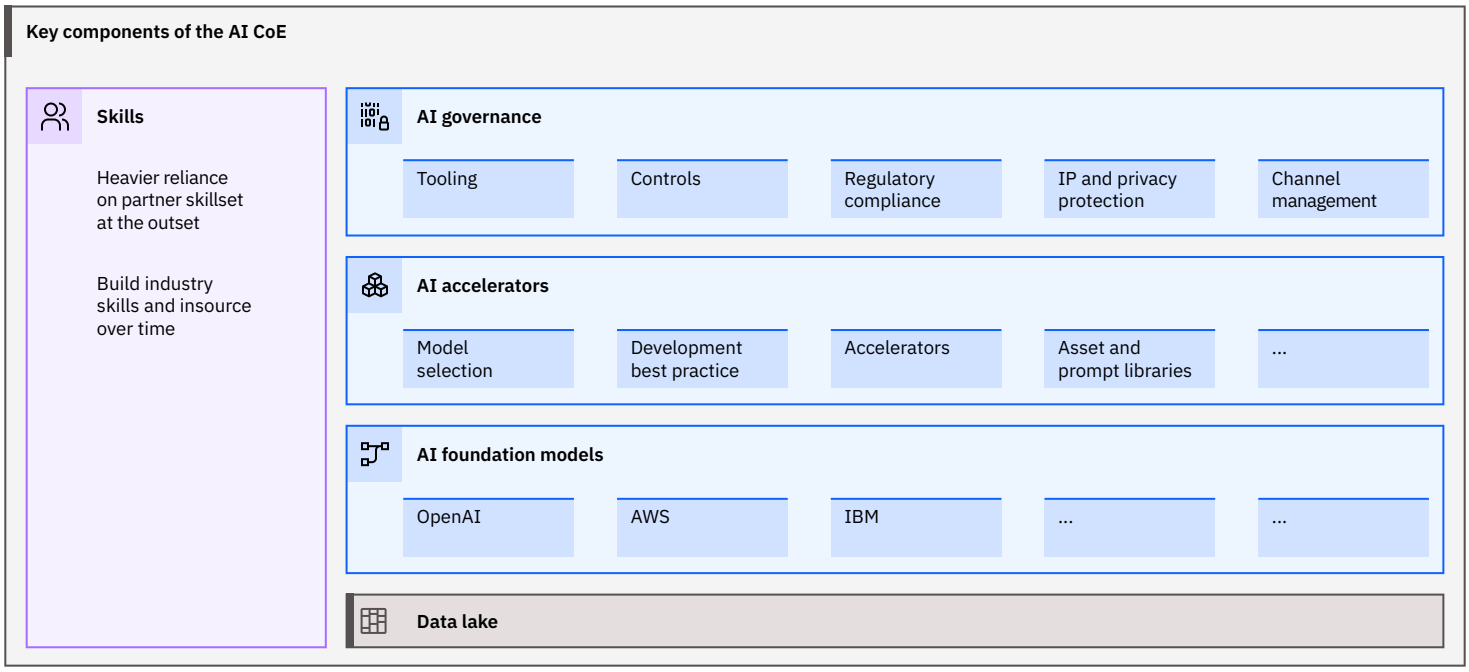

Figure 4. AI CoE supporting a variety of use cases, such as customer engagement, HR, risk and more

Skills and experiential learning

Generative AI is all about people, so automotive companies need to make people, not technology, central to their AI strategy.At IBM, we ran a one-month challenge to train our people hands-on with generative AI tools. This training was based on our core belief in the power of a "show, don't tell" approach to getting a community to use this technology. Putting this technology into the hands of users across all functions and lines of business—not just technology users—allows everyone to understand how transformative it can be to their role and organizational workflows. This approach also provides the critical advantage of getting early buy-in from colleagues in the legal and risk departments. It shows them what the technology is, and more importantly, what it's not. With this knowledge, they can regulate and govern the organization with their eyes open to the opportunities while being mindful of any risks. To that end, and as evidence of this belief, IBM has recently committed to skilling 2 million people on generative AI globally. Our typical approach with clients is for them to lean on our expertise as they start to scale. This enables us to bring the lessons learned from elsewhere—and indeed ourselves—while simultaneously upskilling the client team quickly so they can be self-sufficient as soon as practicable. This approach applies across the enterprise in every level and in every function to gain a full understanding of how to embed this transformative technology.

Governance

Governance is crucial for ensuring responsible and ethical AI use in the automotive industry, especially as AI adoption increases. Generative AI introduces new risks that require cross-enterprise team collaboration for mitigation. A successful enterprise AI governance program requires a multifaceted approach that addresses both organizational and automated AI governance.

The most critical part of organizational AI governance is creating an operating model which involves a variety of steps.

- Establish a clear vision and strategy for AI adoption and governance.
- Assess the current state of AI adoption and governance within the organization.
- Develop a comprehensive framework for AI governance that aligns with industry best practices and regulatory requirements.
- Establish an AI governance committee or office to oversee implementation and ensure compliance.
- Monitor and audit AI systems to ensure they operate within defined parameters and guidelines.
- Establish a process for reporting and addressing ethical or legal concerns related to AI use.
- Continuously monitor and evaluate the effectiveness of the AI governance operating model and make adjustments as needed.

Automated AI governance tools such as IBM watsonx.governance™ are software solutions that help organizations ensure responsible and ethical use of AI by monitoring and managing AI systems throughout their entire lifecycle. These tools are designed to help organizations comply with regulations and industry standards. They also help organizations manage a model's lifecycle and ensure optimal performance and risk management.

- **Compliance:** Automated AI governance tools help organizations comply with regulations and industry standards related to AI use, such as General Data Protection Regulation (GDPR), California Consumer Protection Act (CCPA) and the EU's AI Act. These tools monitor AI systems to ensure they are used in accordance with ethical and legal guidelines and provide features such as data privacy, security and transparency. By using automated governance tools, organizations can demonstrate their commitment to ethical AI use and build trust with stakeholders.
- **Model lifecycle governance:** Automated AI governance tools help manage the entire lifecycle of AI models from development to deployment and maintenance. They provide features such as version control, testing and validation. These features ensure AI models are developed, deployed and updated in a controlled and transparent manner. By using these tools, organizations can ensure their AI models are reliable, consistent and compliant with regulatory requirements.
- **Performance and risk management:** Automated AI governance tools help organizations monitor and analyze the performance of AI models, identify potential biases and inaccuracies, and mitigate risks associated with AI use. These tools provide features such as performance metrics, anomaly detection and risk assessment, which enable organizations to optimize AI model performance and proactively mitigate risks. By using automated governance tools, organizations can ensure that their AI systems are performing optimally and that potential risks are identified and mitigated proactively.

In summary, automated AI governance tools are essential to help organizations that use AI comply with regulations, manage AI model lifecycle and optimize performance and risk management. By using these tools, organizations can demonstrate their commitment to ethical AI use, build trust with stakeholders and ensure their AI systems are reliable, consistent and compliant with regulatory requirements.

Accelerators

The intersection between the AI model and the use cases built on them is of intense interest. We are seeing an increasing trend to abstract away from the model itself to give the end users a set of assets that act as prebuilt assistants and agents for any given task. IBM has an asset in IBM Consulting® called Advantage AI, which helps democratize generative AI. It enables developers, business analysts, UX researchers and architects to collaborate and create reusable prompt libraries and specialized virtual agents and assistants for their on-the-job tasks. Additionally, IBM Consulting is investing heavily in specialized agents such as agents for HR, procurement and customer service. In each case, the agent is a digital product that contextualizes the specific domain it's being instantiated in. This removes the complexity of building these capabilities from the business users themselves.

Foundation models

We believe as multi-model AI strategy is the best option for navigating the diverse landscape of use cases in the automotive industry. This strategy avoids defaulting to a singular model across all scenarios. Instead, it encourages a methodical evaluation of a range of models tailored to specific use-case categories. Factors such as data privacy, data locality, regulations and model interpretability are pivotal considerations in this assessment. Embracing a multi-model framework signifies a deliberate selection of methodologies. It acknowledges that in the dynamic realm of AI, steadfast commitment to a single model approach or technology will limit adaptability to evolving capabilities and industry paradigms. Taking a technology that was developed on the public cloud to a highly regulated industry requires balancing competing trade-offs. These trade-offs include regulatory exposure, optimization for price and performance, copyright protection relating to training data, and the use of the most appropriate computing platform.

However, with a purposeful multi-model approach, automotive companies can be focused and selective about which use cases reside in which environments and utilize appropriate generative AI models, depending on the use case sensitivity. Models are iterating and improving all the time, and confidence in them is increasing. For example, IBM has released IBM watsonx Granite™, a model that is fully indemnified. IBM has listed exactly what data the model has been trained on. The model specializes in enterprise knowledge over general knowledge and is built for a purpose. Therefore, it's smaller than other general-purpose models and, as such, can be hosted where you want it to be, including on premises.

Have your model meet you where your data is and have a model that is mindful of how sensitive the use case is—regulators won't let you have it any other way.

Data lake

As discussed in the foundation models section, generative AI does not fix all known data issues. LLMs are voracious consumers of data, and ensuring the quality of that data is paramount. Often, to feed LLMs, we are rapidly creating a new fused data environment of structured and unstructured data. Typically, we are not able to do that fast enough using traditional techniques. Applying generative AI to the traditional data pipeline process can help reduce cycle time, cut costs and improve data quality. Generative AI tools help identify gaps or inconsistencies and transform data into common structures across the enterprise. We recommend building your generative AI capability in parallel and, more importantly, not having it as a finish-to-start dependency. We believe this is key to making progress while ensuring your data estate is fit for the future.

Foundation models and multi-model strategies

In late 2022, a foundation model took the world by storm. Hundreds of millions of consumers began using ChatGPT to do a variety of tasks, and the choice in the market was simple. Soon thereafter, Google, AWS and IBM announced and introduced foundation models for business. In the next 6–9 months, the marketplace was flooded with different models. Major enterprise software companies, including SAP, Oracle, Salesforce and Adobe, and hyperscalers, including AWS, Microsoft and Google, announced LLMs. Then, fed by venture capital moonshots, many open-source platforms started releasing models, leading to a plethora of LLMs today—likely over one million. As users embraced LLMs, many companies began to introduce small language models (SLMs). These purpose-built models are designed to answer a specific set of questions, unlike LLMs, which are expected to know and answer all kinds of questions. These SLMs are perfect places for deep industry-specific or process-specific content.

At this point, the question to ask is "So what?" What does this dizzying array of LLMs and SLMs mean to the automotive company? Below are some suggestions we offer and follow.

Plan for a multi-model strategy
A multi-model AI strategy is the best option for navigating the diverse landscape of companies, LLMs and SLMs. This strategy avoids defaulting to a singular model across all scenarios. Instead, it encourages a methodical evaluation of a range of models tailored to specific use case categories and nonfunctional requirements.

Recognize that it will be about efficacy
If the purpose of generative AI, LLMs and SLMS is to augment human capabilities, then we need models that produce outputs that have an extremely high degree of efficacy. Here are three observations about efficacy.

1. Efficacy will depend on which activity or process you are attempting to perform, such as project, business, IT, marketing, service or process.
2. Much of the efficacy we are achieving today is not based on prompt engineering but a framework of data ingestion, retrieval-augmented generation (RAG) patterns, vectors and embeddings that are all designed to provide the models with context.
3. The entire framework needs to be elastic enough for the arrival of the next greatest model.

Plan to be private when you need to be
We have observed that clients in regulated industries with older legacy systems have trade secrets or IP embedded in most of their core systems. Many of our clients do not want to expose their IP or trade secrets in the public cloud, so they demand that all models or frameworks be able to run locally. Additionally, data privacy, data locality, regulations and model interpretability are pivotal considerations in this assessment.

It's still all about the data
Just as generative AI came into the limelight, the world thought it would fix all data issues. Retrieving data from anywhere, preparing any type of data and fixing data errors—generative AI was going to solve it all. The first and most important question to ask about LLMs is: where did the data come from to train them? Court dockets are full of model providers that sourced data for one usage pattern and then turned it public. IBM has taken careful steps to source all training data appropriately and work with model providers who have done the same. But that is not enough. The earlier section on efficacy discusses a framework to ingest vast amounts of data into models. The framework increases efficacy by providing context to the models. That entire pipeline of data is different from what most clients have dealt with over the past 30 years. It's mandatory for any model provider you work with to have a rock-solid framework for augmenting the capabilities of the model in a safe and secure manner.

05

The IBM approach to generative AI-based transformation

IBM is making material investments in generative AI across the enterprise. Furthermore, we are applying these investments to what we do internally to both accelerate our own transformation and take the lessons learned to our clients to unlock value for them faster.

Thought leadership
We can help you understand the opportunity and identify the actions to take: the "how", not just the "what". The IBV thought leadership unit publishes key reports and action guides to help enterprises plot a path forward around generative AI, including AI ethics guides from our AI ethics board. In addition, the IBV team has published benchmarking data that reflects the impact of generative AI on most enterprise processes. The IBV paper, "Generative AI: State of the Enterprise Market: Automotive" highlights that automotive companies have been somewhat slower to start on generative AI than other industries. It also indicates that 42% of automotive executives report that their companies are not just evaluating generative AI but are in some phase of adopting it, including piloting, implementing, operating or optimizing it.[3] Use cases are emerging in all areas, but the largest number of active projects can be found in information security and IT, followed by product development and risk and compliance.

Research
Our IBM Research experts have been pioneering work in domain-specific foundation models. We have built and released a number of models, with more on the way. These models include LLMs that are free from copyright and IP infringements and where we provide indemnification. We have models that support the use of generative AI in forecasting and decision-making. Additionally, we have models and new algorithms in specific areas, such as those based on time-series data, to improve the prediction of operational issues.

Technology
Our industry-leading technology platform for generative AI, IBM watsonx™, has been designed to help enterprises adopt and scale generative AI safely and securely within complex, regulated environments. It has an end-to-end process for building and testing foundation models and generative AI. The process starts with data collection and ends in control points for tracking the responsible deployment of models and applications, and focuses on governance, risk assessment, bias mitigation and compliance. The watsonx platform uses filtering and scoring technologies to reduce the likelihood of including copyrighted material in a system's output. IBM also applies rigor to the development and testing of its foundation models, with contractual protections for IBM-developed generative AI products and models.

As a means for democratizing LLM development in watsonx.ai™, IBM Research and Red Hat® created InstructLab (Large-scale Alignment for chatBots), an open-source initiative that encourages community-driven enhancements to promote rapid model development through collaborative contributions. The associated toolkit systematically generates synthetic data for chatbot tasks and for assimilating new knowledge and capabilities into the foundation model—without overwriting what the model has already learned. Through the InstructLab project, IBM and Red Hat have released select open-source licensed IBM Granite language and code models under the Apache 2.0 license.

Consulting
The technology choices you make are crucial, but so is having a partner with the expertise to put that technology to work. Our IBM Consulting team of over 160,000 team members deeply believes in the power of an ecosystem-led approach. Our team of over 21,000 data scientists, AI engineers and AI consultants have worked on thousands of engagements with clients. We have helped them achieve transformational value from AI technologies in their organizations. We are also a leading partner to AWS and Microsoft with, respectively, over 10,000 certified consultants and 38,000 specialist practitioners. We can provide in-depth, strategic business and technology expertise locally to your teams and use the scaled power of our delivery center network of over 70 locations. We can help our clients build their own in-house AI capabilities and augment anything that is missing.

Assets and accelerators
Based on our experience across thousands of other AI engagements, we have developed a toolkit of assets and accelerators to help accelerate your value realization. For example, we have a LLM and framework to support the use of generative AI for the software development lifecycle. This library of reusable assistants—purpose-built for specific steps in the software development process—can help developers increase their productivity by 20%–30% across the board. Additionally, developers can be 60% more productive in select areas such as test case generation.

Ecosystem partner
Across all areas of IBM, we follow an ecosystem-first approach. We have deep and established partnerships with cloud providers, such as AWS and Microsoft, and rank as a leading service provider for clients on their technology. Over 11,000 of our consultants work with leading automotive companies across the globe. They help companies implement and get the most out of enterprise applications, such as Adobe, Salesforce, SAP, Oracle, Workday, ServiceNow and more. We are committed to open standards and are partnering with open AI firms to enable you to take advantage of open-source innovation in both a technology and industry context. One such company is Hugging Face, an open-source model hub that is integrated into our IBM watsonx.ai AI studio.

IBM as a user of AI: Our own learned experience of scaling AI
We have been on a multiyear journey to infuse AI and generative AI into our organization at scale. As a result, we can apply the learnings, assets and accelerators from this work to help the automotive industry. For example, over 94% of all employee-HR interactions at IBM are now carried out without human intervention using AskHR, a conversational AI-based chatbot. It has transformed the employee experience with HR and delivered significant productivity enhancements. In addition, our recent company-wide challenge focused on generative AI to facilitate engagement across the company. The intention was to help our employees develop generative AI skills and spark their curiosity for applying the technology to their areas of work. We had more than 158,000 people participate, with over 23,000 teams organically forming across the company and over 12,000 ideas and submissions reviewed by a judging panel over the course of several weeks. We believe that unlocking this type of approach to driving the skills, learning and change agenda within your organization is critical.

06

Implications for the automotive industry and how IBM can help

There are 3 key takeaways automotive companies should focus on to enable a sustained and tangible ROI from investing in AI.

1. Center efforts around a business case with metrics
It's easy to waste a significant amount of time and money chasing technology proof points. We recommend a fully vetted enterprise-wide change program with people, process and technology metrics that track value across the enterprise.

2. Establish an open foundation
To drive transformative change, you will be operating in a multi-model world with a strong underlying framework to drive the efficacy of model output. For the past two decades, enterprises have spent millions of dollars to modernize systems that were not designed to be open from the get-go. Why start your generative AI journey with a closed environment? We strongly recommend you start with an open environment that provides flexibility. Embedding the right guardrails, control patterns and adoption frameworks in that open environment will be key to sustaining and accelerating the ROI from your investment in AI.

3. Equip your people
We have found that delivering generative AI at scale requires three levels of skill: foundational, experienced and expert. Rather than discussing the ratios of those skill sets, let's focus on their capabilities. For foundational skills, we expect every one of our consultants to be proficient in generative AI, enough to use prompts to accelerate their work. For experienced skills, we expect our consultants to be able to discuss a variety of generative AI tools with clients and how to use them to generate value. They also need to be able to write complex prompts across multiple business processes. The expert capability is reserved for deep data engineers and data scientists, who will be driving model efficacy using the framework described in this document.

You should establish a detailed training plan with metrics and measurements that drive into people's annual objectives.

In closing, we would like to suggest some immediate areas where we believe we can support you in the next steps of your generative AI strategy and execution plan.

Building tomorrow's strategy while realizing value today
IBM Consulting can work with you on your strategy, execution and roadmap for scaling generative AI. They happen alongside quick wins that can unlock efficiencies in areas such as IT, software development, agent assistance in multiple domains, support functions such as HR and procurement, dealer management and more. Here, we have direct experience, assets and accelerators that will help you achieve outcomes more quickly.

Skilling your organization to scale generative AI throughout the enterprise
IBM Consulting can help you lay out and execute a plan to upskill your entire team on generative AI. This process entails using both governance structures and methods, and upskilling your workforce by training them hands-on and developing their skills in a safe and secure manner.

Controlling AI with enterprise-scale governance and tooling
IBM Technology can help you understand the architectural footprint required to seize the opportunities that come with generative AI and invest to build a scaled "air traffic control" or ecosystem model. With our help, you will have the right guardrails and controls to support a multi-model approach to implementing generative AI within the enterprise. Additionally, we can implement our watsonx.governance tool to act as an orchestration platform across various models and technologies.

Generative AI offers an incredible opportunity for virtually every enterprise globally. Automotive companies are particularly well-positioned to take advantage of this exciting new technology and we look forward to partnering with you on your journey.

Case study

Auditing dealer websites
In most countries, automotive companies sell their vehicles through a network of dealers. Each dealer's online presence is independent, although typically supported by the OEM's marketing organization. The challenge for the OEM is that dealer websites have a significant variance in how well they follow the guidance provided by the OEM. Deviations from that guidance can lead to law enforcement actions against dealers, as well as damaging consumer sentiment about dealers and the OEM.

The IBM team partnered with Adobe to develop a generative AI-powered compliance dealer website audit workflow for a global automotive OEM. The workflow is defined and automated in Adobe Workfront and the actual auditing of website content is done with IBM watsonx generative AI. The first release of the solution audits websites from three perspectives:

1. Brand representation: Checks for correct trademark and logo images and placements. Validates business taxonomy (terms such as MSRP, ABS, PHEV, MPG and more).
2. Legal exposure: Validates promotion details and financing or leasing offers such as interest rates, discounts and promotions. Checks regulatory compliance such as privacy statements, legal statements and disclaimers.
3. Client experience: Checks for security and vulnerability, content tone to ensure professional or neutral tone, and HTML formatting and hyperlink validity.

The solution improves the OEM's auditing capability in two ways. First, the automated capability allows the OEM to audit all dealers' website content instead of the sampling approach that they were forced to take when the auditing was done manually. Second, it significantly improves the experience of the human auditors. Instead of doing the relatively dull task of crawling through websites, the auditors can now focus on the high-value task of resolving issues identified by the generative AI workflow.

1. Rebuilding Trust in 2024: Davos in summary, TechInformed, January 2024
2. IBM Institute for Business Value, “Generative AI: The state of the market—automotive,” 14 July 2023
3. IBM Institute for Business Value, “The CEO’s Guide to Generative AI: Marketing”, December 2023

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
July 2024