AWS FOR ENERGY AND UTILITIES

The transformative power of generative AI in the energy industry

Table of contents

Executive Summary

Generative AI is poised to revolutionize the energy industry, offering unprecedented capabilities in decision-making and problem-solving. Energy companies are currently focused on augmenting human capabilities, but this approach falls short of generative AI's full potential. Six critical energy industry domains will be transformed: Subsurface, Midstream, Downstream, Trading, Decarbonization, and Grid Operations.

1. For subsurface oil and gas, AI agents will streamline workflows, enhancing exploration and production efficiency.

2. Midstream operations will benefit from AI-driven pipeline integrity management, transportation optimization, and storage management.

3. In downstream operations, generative AI will optimize refinery processes, improve safety, and increase operational efficiency.

4. Energy trading will see improvements in deal booking, risk management, compliance monitoring, and financial settlements.

5. Generative AI will accelerate decarbonization efforts through innovations in carbon capture, material science, and clean energy production.

6. In grid operations, generative AI can optimize real-time decisions, improve system reliability, and enhance customer experiences.

The adoption of generative AI will lead to significant productivity gains and cost reductions across the energy sector. The competitive position of energy companies will increasingly depend on the availability of high-quality data and deployments of workflows enabled by sophisticated AI models. Generative AI will enable more autonomous and efficient operations, shifting human roles towards strategic decision-making and oversight.

Implementing generative AI requires careful consideration of data quality, model selection, and integration with existing systems. Energy companies should prepare their data, employees, and mechanisms to harness the transformative power of generative AI. In parallel, companies must address challenges such as hallucinations, responsible use, transparency, and security in AI implementation. Collaboration between regulators and companies is necessary to balance AI exploration with these safety concerns.

The energy industry stands at the cusp of a technological revolution that will redefine how it addresses global energy needs and environmental concerns.

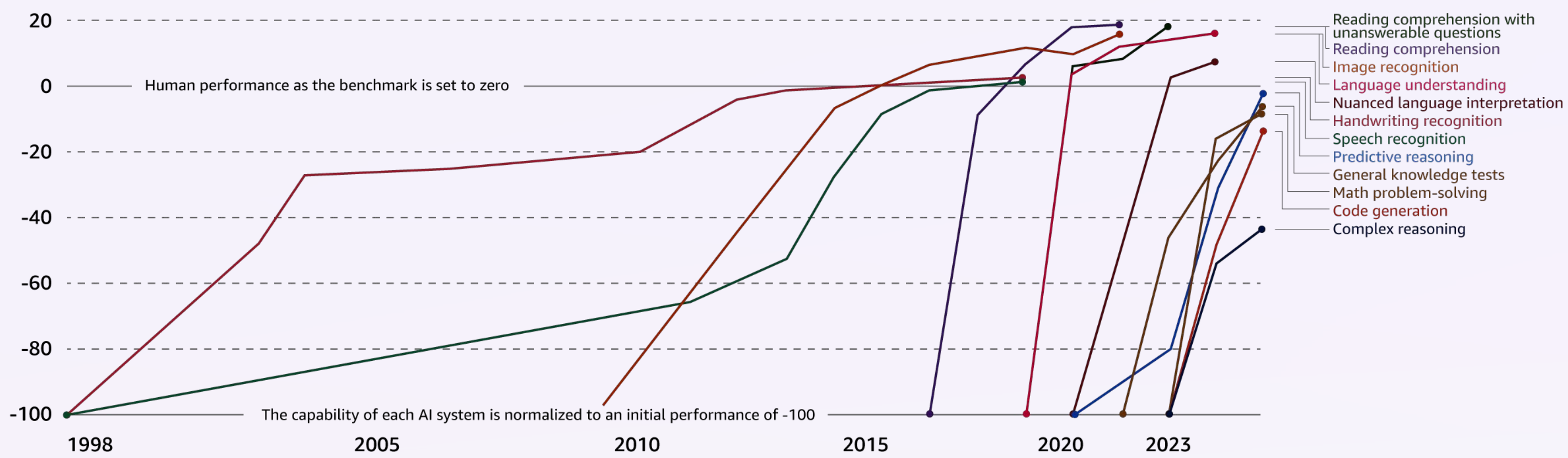
The transformative power of generative AI in the energy industry

Generative AI is already transforming the world. Our ability to make decisions and predictions accounting for every piece of science, research, conversation, learnings, and any other information ever created by humanity is providing a power we are only now beginning to understand and harness. Those who adopt this power first can disrupt their industries to unprecedented levels. The Energy industry is no different and that is a good thing: doubling energy production while reducing global greenhouse gas emissions to hold the global temperature increase below 2°C above pre-industrial levels by 2050, is a major engineering, regulatory, and social challenge that needs new materials and engineering, economic models, and trillions of capital investments. Generative AI can help solve and accelerate the solutions to these challenges. In the last couple years, businesses have allocated considerable effort to learn and experiment with generative AI technology and to develop internal mechanisms to adopt it safely. Although we have completed dozens of use case automations in subsurface, trading, customer care, and supply chain to name a few, there is more that needs to be done to prepare for the disruptive power that is coming in the next two to three years. Some executives question to what extent generative AI can be disruptive, and they may be tempted to take a ”wait and see” approach. That could be a costly, and perhaps fatal, mistake.

While we can’t predict the future, extrapolating current patterns in deep learning can help us understand what levels of capability generative AI can attain in the short- and -term future. The projections are nothing short of astounding. After all, just 10 years ago, AI underperformed humans in language and image recognition in every test. Today, they beat humans in many tests and are getting closer in all of them. This graph from Kiela et al. (2023), with minor processing by Our World in Data, shows the progress over the last two decades.

Test scores of AI systems on various capabilities relative to human performance

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI’s performance crosses the zero line, it scored more points than the humans.



Data source: Kiela et al. (2023)
Note: For each capabilitiy, the first year always shows a baseline of -100, even if better performance was recorded later that year

In 2019, leading Large Language Models (LLMs) attained preschooler reasoning and could answer questions with proper grammar. In 2020, LLMs ”graduated” to elementary schooler reasoning and could string paragraphs together. In 2023, they passed “high schooler” level and could reason and create intelligently. There are multiple compounding trends that can further accelerate this progress: 1/the advances in compute coming from Moore’s law but also improvements in machine learning (ML) specialized chips such as GPUs and TPUs, 2/programming algorithms and paradigms such as Mixture of Experts, 3/model enablement and ecosystem advancements such as scaffolding, chain of thought, and agents, and 4/the billions if not trillions of dollars pouring into the field. Artificial General Intelligence (AGI) is likely around the corner and commercially available by the end of the decade in advanced LLMs. By the end of decade, there will be fleets of thousands of LLMs with superhuman reasoning abilities that can take into account every piece of data ever created with thousands of parameters and dependencies interacting with each other. This will be far beyond the capabilities of any human team. Perhaps the most important moment to look for is when AI systems conduct advanced AI research and independently incorporate improvements on their own. This will unleash an unprecedented positive feedback loop where AI improves itself and will represent an inflection point in terms of the velocity at which AI will progress subsequently.

Most Energy companies are focusing their generative AI efforts today on augmenting humans by lightening their cognitive load or automating existing processes that need quick retrieval and summarization of information for repetitive, low-risk processes such as internal helpdesks. Although those are good early steps, they fall short of current and quickly approaching generative and general AI capabilities by taking an ”incrementalist” approach. They fall short of incorporating the full capabilities of generative AI to maximize business impact. For example, generative AI-based workflows should include agentic collaboration where specialized AI models take on a workflow task that they have been specifically trained for. With this approach, the team of agents employed throughout the workflow, significantly enhances their collective problem-solving capabilities when compared to a general foundational model. A good example would be a subsurface workflow with agents that acquire and prepare data, agents that define pre-models, agents that perform subsurface characterization, and agents that plan reservoir exploitation activities.

One more capability quickly gaining traction is foundation model (FM) self-inspection. With self-inspection, a specialized AI model can inspect and correct its output in order to increase response reliability. Beyond stitching AI models in an agentic workflow, modern generative AI models can also use tools such as accessing web pages or executables resulting in material increases to productivity and expansion of problems that can be addressed.

With these competences, the workflows addressable by generative AI are broader than the ones most companies are currently tackling. For example, if a piece of rotating equipment fails, there are usually too many relevant questions to ask to get to the root cause: who worked on that equipment, what procedures were followed, what happened during the last maintenance, what were the conditions at the point of failure, which supplier provided the parts or services, and did those suppliers have other issues in the past? A human with enough time could answer these questions, but with generative AI, we can answer those questions in minutes. With generative AI, a team of agents can gather relevant data such as model, condition failure, and the weather from internal and external sources, an engineering trained FM can triage the failure mode, and an Operations FM can plan the technicians, parts, and ideal timing of a repair based on supply chain and market conditions. An LLM specialized in performing the operational root cause analysis can update any maintenance procedures automatically if applicable.

This narrative explores how generative AI will transform six critical energy industry domains: Subsurface, Midstream, Downstream, Trading, Decarbonization, and Grid Modernization. We also explore the opportunity for companies to think bigger about their data and considerations that are critical for energy companies to harness this unprecedented power and prepare their data, employees, and mechanisms for it today.

1. Generative AI in grid and utility operations

The power grid is one of the most complex systems on earth. It is facing new challenges from several coinciding trends, such as the growth of inverter-based intermittent/distributed generation resources, the growth of data center loads, the electrification of transportation, the new grid enhancing technologies, and the rising threat of extreme weather events. These challenges raise unprecedented challenges to maintain grid stability, especially during emergency conditions. A generative AI Powered Grid Control Room could make real-time decisions (or in the near-term support the operators) to take optimal control actions with confidence to balance supply and demand, maintain system frequency, and relieve congestions. The generative AI Grid Operator (G2O) can help reduce renewable curtailment by 30% while improving system reliability. The generative AI Grid Operator consists of: 1/ an enhanced grid digital twin with integrated data layers from SCADA, PMU, IoT, GIS, Weather, and so on; 2/ a physics-informed grid deep learning model, continuously trained with real-time grid data and many “what-if” grid simulations, using state-of-the-art grid analytical tools powered by large High Performance Computing (HPC) clusters; 3/ an LLM that is fine-tuned with grid operational protocols, grid common information model (CIM), and historical operator event log; 4/ a comprehensive system impact test with real-time grid simulation to make sure of the feasibility and safety of recommended operator actions; 5/ an after-fact “perfect dispatch” tool to simulate the best possible control action if G2O had a perfect prediction of the system. “Perfect dispatch” is used to rank G2O actions and to improve the accuracy of Grid LLMs, creating a self-improving feedback loop that takes into account every condition in the grid; 6/ a user interface for operator inspection and control. The G2O will be physics based, reliable, responsible, secure, transparent, and explainable. We recommend implementing it first in an operator training environment to earn trust from operators and decision makers, and to inform better dispatch and grid operations while training the G2O. Taking an example of a storm hitting part of the grid, the G2O will overlay the weather map onto grid topology, and conduct “N-K” contingency analysis for an impacted portion of the grid. Then, it will use ML to generate outage prediction and identify potentially harmful anomalies. The physics-informed grid deep neural network model can quickly produce sets of preventive controls to minimize the system impact. The G2O will feed the set of preventive controls into the Grid LLM to produce the best actions, based on grid operational protocols and ranked historical operator actions.

In addition to transforming Grid operations, generative AI will also change the customer experience of utility customers. Generative AI is already transforming contact centers across almost every industry by enabling businesses to deliver exceptional customer experiences, improve operational efficiency, and gain valuable insights from customer interactions. Generative AI can be applied to the entire customer interaction process, from Pre-Contact Insights and Automation (proactive intent identification and predictive routing), to During-Contact Agent Assistance (customer interaction summary, personalized real-time recommendation, making sure of compliance during conversation, and call/chat sentiment analysis), and to Post-Contact Optimization (Insights and call/chat summary, automatic ticket/incident creation, and agent training). Generative AI powered contact centers will not only help energy retailers and distribution network operators (DNO) provide superb customer services, but also enable more customers to actively participate in grid management and the energy transition. This needs a generative AI contact center to integrate with various enterprise data sources, both structured and unstructured, which will improve the responsiveness and accuracy of fine-tuned LLMs and reduce hallucinations. Furthermore, Retrieval-Augmented Generation (RAG) gives an LLM access to customer data such as meter data, and billing information, while making sure of customer privacy and security. For example, a customer could send a photo of a problematic bill and generative AI could understand its key elements (for example bill date, bill items), gathering corresponding meter data from a Meter Data Management platform, and immediately providing contextual help to the customer. During the contact, generative AI will identify cross-sell/up-sell opportunities and automate the buying journey for EV Chargers, Solar, or Battery Storage. Generative AI could use the information from OMS, DMS, GIS, MDM, and ERP to map customers to transformers/feeders, dispatch crews for repair, and inform customers with the cause of outage and the estimated restoration time. When the system is anticipating reserve scarcity/emergency conditions, generative AI could use real-time data from ADMS, and DERMS to proactively contact customers with personalized recommendations/control, such as thermostat settings and EV charging schedules, to encourage customers to help balance the grid.

2. The future of generative AI in subsurface oil and gas

Energy companies are grappling with the challenge of meeting rising global energy demands while making sure of safety, minimizing environmental impact, and maintaining affordability. Historically, the energy industry has adeptly addressed growing demands through ground-breaking explorations such as 3D seismic imaging and advanced drilling techniques. However, this progress has also resulted in an immense accumulation of data exceeding the human capacity for analysis.

To address this, many energy companies are turning to generative AI for rapid data retrieval and enhanced productivity through advanced chatbots. Although LLMs are notable for their text generation, sentiment analysis, and code creation capabilities, their generalist design can fall short when handling the specialized language and data unique to the energy sector.

To overcome these limitations, companies are using tools such as Amazon Bedrock to select and customize high-performing FMs tailored to industry-specific needs. This fine-tuning with proprietary data enables companies to develop their own domain-specific AI with insights and expertise that surpass that of any human expert.

FMs are evolving from assistants to near-autonomous agents, and can run tasks and adapt their responses based on external inputs and critiques from other agents. These FM-agents will seamlessly integrate internal and external tools through APIs, filling knowledge gaps and coordinating workflows. This evolution creates a virtual team of experts akin to having hundreds of PhD-level specialists that can interpret vast amounts of data and performing complex tasks. This advancement will drive unprecedented improvements across the energy sector, enhancing everything from seismic interpretation and well planning to resource assessments and economic evaluations.

For example, an exploration geologist seeking to identify a drillable prospect could deploy an FM-agent to identify, select, and visualize data in a specific area of interest. Then, the agent can interpret the seismic data using its extensive knowledge of structural and stratigraphic geology, and seismic processing combined with access to the prior seismic interpretations from the organization. Then, the geologist can instruct the FM-agent to review the leads, prospects, drilled targets, and analogues to generate a comprehensive hydrocarbon volume assessment and chance of success. This iterative process and dialogue with the FM-agent allows the geologist to continuously refine and validate insights.

Similarly, a well engineer can use an FM-agent to review the nearby drilled wells and generate a detailed well plan, significantly boosting efficiency, accuracy, and safety. With just a few clicks, both the geologist and the well engineer can harness the FM-agent's vast database and analytical capabilities to optimize well placement and drilling operations in methods previously unimaginable. The FM-agent not only streamlines workflows but also enhances decision-making, leading to more informed, faster, and explorative approaches for the energy sector.

To fully harness the power of FM-agents, companies must first address the critical issue of data quality and availability: incomplete or inaccurate data results in ineffective AI. Energy Data Insights on Amazon Web Services (AWS) addresses this challenge by providing energy companies with a fully managed OSDU® standardized data platform, making sure of data consistency, quality, and interoperability across the industry. Without this robust data foundation, the effectiveness of FM-agents and the insights they provide would be significantly compromised, thus limiting their potential.

For human subsurface asset teams and drilling engineers, this transformation shifts their role from routine data gathering and analysis to strategic decision-making, oversight, and creativity. With AI handling data interpretation and analysis, oil and gas professionals can focus on applying and evaluating AI-generated insights and result implications, using their expertise to improve accuracy and drive more strategic and explorative approaches for their organization.

Looking ahead, the competitive landscape will increasingly hinge on both high-quality data availability and the sophistication of FM technology. Companies that excel in harnessing a robust data foundation and deploying advanced FM-agents customized to their data will set new industry standards and gain a decisive competitive edge. The future will see a race where success is defined by gathering high-quality data and having the most advanced foundational models, setting benchmarks for operational excellence and exploration success rates.

3. New paradigm in downstream

The downstream oil and gas sector employs two key forecasting scenarios models: 1/ Demand-driven: runs forecast based on month-on-month change in domestic consumption of fuel products, and 2/ Capacity-driven: runs forecast based on capacity projections, historical patterns in usage, and seasonal maintenance, without considering the impact from domestic or internal fuel product demand. The global downstream investments are estimated at $611 billion in 2024, up 7.2% from 2023. These investments primarily target 1/ lower cost of supply, 2/ capturing value from wellhead to market, 3/ improving portfolio value, such as blue fuel products, and 4/ leading in sustainability, reducing GHG emissions and increasing circularity. According to Accenture's "The Intelligent Refinery" report, 80% of refiners reported that digital is adding up to $50 million in value to their business, and 75% stated that they intend to spend more on digital over the next three to five years. However, there is potential for orders of magnitude increases in value from digital improvements with AI and generative AI.

Digital transformed refineries can: 1/ improve net cash margins. Refinery operations traditionally involve a variety of complex optimization algorithms across feedstock, plant scheduling, blending, and distribution chain to maximize net cash margin. These algorithms are not optimally exploited currently, but can be with AI and GenAI; 2/ Be safer, more reliable, and more energy-efficient. Generative AI can analyze historical data, simulate scenarios, and provide insights for predictive and prescriptive maintenance, reducing downtime and reducing energy intensity; 3/ Use AI-enabled solutions such as generative AI. Generative AI can be used for tasks such as generating synthetic data for training ML models (for example corrosion monitoring), creating virtual environments for testing and simulation, deciphering molecular chemistry, and generating natural language reports and insights; 4/ Be more attractive as a workplace for top talent.

The transformation begins with connecting and collecting data from SCADA systems and IoT sensors across the refinery, capturing real-time data into processes and equipment performance. Then, this data, combined with synthetic data generated from simulations and generative AI models, is used to create integrated and contextualized geospatial digital twins, virtual replicas of physical assets and processes. These digital twins enable the simulation and testing of various scenarios without disrupting actual operations, allowing for risk-free experimentation and optimization. ML algorithms are applied to the vast amounts of data collected, enabling predictive maintenance, process optimization, operational efficiency improvements, and streamlining supply chain related activities, such as improving demand-capacity planning forecast accuracy. Then the ML will create a commercially optimized of unit set points, and a blend plan that is constrained by safety and physics/chemistry.Those are pushed back into production, closing the loop by informing real-time decision-making and automating processes. This continuous cycle of data collection and contextualization, digital twin simulation, ML-driven optimization, and implementation drives ongoing improvements in refinery operations, reducing downtime, optimizing yield, and enhancing safety and sustainability.

4. Midstream

The midstream oil and gas sector, which encompasses the transportation, processing, and storage of hydrocarbon molecules, has seen an increase in pipeline and storage capex of 7%, which is $105 billion in 2024 when compared to $98 billion in 2023. These investments primarily target pipeline expansion on 1/ onshore assets driven by an increase of Gas-Oil-Ratio (GOR) and new drilling activity with long laterals in tight formations, 2/ offshore assets driven by a boost in infrastructure-led-drilling activity and enhanced-oil-recovery methods, and 3/ LNG infrastructure, liquefaction capacity advances with the expectation of an increase in gas (transition fuel) consumption.

This sector's strategic priorities can be anchored on 1/ maximizing the value of the assets, using intelligent decision-making in day-to-day operations by connecting demand/supply outlook, variation in prices, and disruptions. Proactively identifying, assessing, and mitigating supply chain risks through advanced modeling simulations. Providing line of sight to disruptions and supporting decision-making to identify alternate capacity options. 2/ Infrastructure project execution on-time and on-budget by using engineering workflow automation and real-time collaboration between n-tier suppliers. And 3/ rapid time-to-value for operation performance improvement and route planning.

The generative AI capabilities can significantly transform the midstream strategic priorities, such as the following:

- Pipeline integrity by analyzing vast amounts of data from inline inspection tools, manufacturer and research documentation, sensors and drones, historical weather, incident patterns, and other public and private information to assess the risk of pipeline failures and schedule maintenance activities around planned major maintenance events. Generative AI FM will also drive optimization workflows to identify the best course of action and prescribe optimal solutions such as pipeline drag reducing agents.

- Transportation, where generative AI will run hundreds of route combinations and determine the optimal transportation routes for midstream operations by considering factors such as pipeline capacity, product demand, weather, most favorable contracts, scheduling, logistics, and regulatory constraints. This will lead to more efficient and cost-effective transportation of hydrocarbons, while making sure of timely delivery.

- Storage management, where generative AI will analyze historical demand patterns, production forecasts, and market conditions to optimize inventory levels at storage facilities, reducing carrying costs and mitigating supply-demand imbalances. Generative AI will also monitor real-time data from sensors and cameras at storage facilities to detect potential issues, such as leaks, spills, or equipment failures. Based on the issue, safety risk, and availability, it will dispatch the right crews and equipment for fast resolution.

- Compliance, where generative AI will monitor regulatory requirements, industry standards, and environmental regulations, eliminating penalties for operators. Reporting will be filed automatically and complete audit trails of compliance filings kept. In addition to regulatory compliance, generative AI will also make sure of compliance of commercial invoices with sensor and meter data, contracts and amendments, and accounting principles, thus eliminating discrepancies, improving cash flows, and streamlining accounting activities.

5. Generative AI: a model shift in energy trading

The energy trading industry stands on the cusp of a technological revolution, with generative AI poised to transform every facet of operations. From deal booking to financial settlements, this cutting-edge technology promises to streamline processes, enhance decision-making, and mitigate risks in ways previously unimaginable. This impact will be felt in each of the five key Energy Trading domains: Deal Booking and Trade Capture, Risk Management, Compliance Monitoring, Financial Settlements, and Accounting.

- Deal Booking and Trade Capture: Using natural language processing and ML algorithms, AI systems will automatically extract key information from various sources such as emails, chat logs, and voice recordings. This capability will eliminate data entry and reduce processing time. Moreover, AI can generate standardized contract language, making sure of consistency across deals while adapting to specific counterparty requirements. It will learn from historical data to use optimal deal structures, uncovering new trading opportunities that human traders might have overlooked.

- Risk Management: Specialized Foundational Models will analyze all historical and real-time data, enriched by external data such as geopolitics, weather, and historical operations to generate highly accurate price forecasts and risk scenarios. These models can simulate countless market conditions, allowing for automated portfolio stress-testing against a wide range of potential outcomes. Furthermore, FMs will continuously monitor market conditions and automatically adjust risk parameters, making sure that exposure remains within acceptable limits even in rapidly changing markets.

- Compliance Monitoring: Compliance will be assured by continued analysis of trading patterns, communications, and market data in real-time to detect potential regulatory violations or market manipulation. By generating synthetic datasets, AI will train itself on a vast array of compliance scenarios, such as rare or hypothetical situations that may not exist in historical data. This approach will enhance the system's ability to identify subtle compliance issues that might historically escape human detection. Generative AI will also generate detailed compliance reports and explanations, providing regulators and auditors with transparent, data-driven justifications for trading decisions.

- Financial Settlements: Generative AI will automate the reconciliation of trades, invoices, and payments, eliminating the time and resources needed for this traditionally labor-intensive task. By generating detailed settlement reports and identifying discrepancies, generative AI will streamline dispute resolution processes by proactively suggesting solutions, minimizing delays, and improving cash flow management.

- Accounting: Generative AI promises to enhance accuracy, efficiency, and insights in this domain. Generative AI systems will automatically generate journal entries, financial statements, and management reports, making sure of consistency and eliminating human effort. By analyzing historical financial data and market trends, generative AI will generate accurate accruals and provisions, improving the precision of financial reporting. Moreover, generative AI will produce advanced forecasts of future financial performance based on current trading positions and market conditions under different scenarios, facilitating informed strategic decisions.

aws

6. Generative AI for decarbonization: empowering energy companies

The race against climate change has reached a pivotal juncture, demanding disruptive solutions that transcend traditional approaches. As the energy industry grapples with the urgent need for decarbonization, the availability of generative AI presents a paradigm shift that could redefine the very fabric of sustainability efforts. Generative AI has already demonstrated its early prowess in optimizing complex processes, simulating scenarios, and generating actionable insights.

Reimagining Carbon Capture and Storage (CCS): GenAI's ability to iteratively learn, adapt, and evolve could revolutionize CCS processes. Imagine GenAI systems that continuously refine capture techniques, dynamically optimize transportation routes, and autonomously identify novel storage sites tailored to specific geological conditions. This symbiotic fusion of generative AI could transcend the limitations of current CCS approaches, paving the way for a future where carbon capture becomes an integral, self-optimizing component of energy production.

Material Science Breakthroughs for Direct Air Capture: Generative AI could accelerate the discovery of novel materials to revolutionize direct air capture (DAC) and atmospheric carbon extraction. These systems can continuously scan vast databases of compounds, running simulations to identify promising candidates. Then the most viable options can be synthesized and tested, with GenAI autonomously optimizing experiments and analyses. This iterative loop of simulation, testing, and refinement could rapidly uncover explorative sorbents, membranes, and catalysts tailored for carbon capture from ambient air.

For example, generative models could identify novel porous materials or metal-organic frameworks with optimal surface areas, selective absorption, and low regeneration energy penalties. Meanwhile, GenAI-driven autonomous labs could synthesize and validate these compounds. As new materials are proven, GenAI could combine them into modular DAC systems specialized for deployment across various geographies and conditions. It could also integrate them into existing infrastructure, from factory smokestacks to commercial HVAC units, to passively extract carbon. This fusion of computational material discovery and autonomous rapid experimentation could transform carbon extraction technologies, accelerating the march to carbon neutrality.

Carbon Intensity Prediction and Mitigation: By harnessing the power of recursive learning and predictive modeling, GenAI systems could anticipate and mitigate carbon intensity fluctuations before they occur. Models will continuously analyze a multitude of variables, from production processes to supply chain dynamics, and generate proactive strategies to minimize carbon footprints. This paradigm shift would empower energy companies to stay ahead of the curve, proactively addressing environmental concerns and driving sustainable growth.

Autonomous Carbon Credit Ecosystem: Imagine an autonomous ecosystem governed by GenAI, where carbon credits are dynamically managed, traded, and optimized in real-time. Generative AI will seamlessly integrate market data, regulatory frameworks, and project-specific variables to generate predictive models and recommendations for carbon credit portfolios. This self-regulating system could autonomously adapt to market fluctuations, capitalize on emerging opportunities, and make sure that energy companies maximize the value of their carbon credits while contributing to global emission reduction efforts.

Intelligent Clean Energy Production: The convergence of generative AI could propel clean energy production into a new era of efficiency and scalability. GenAI systems could autonomously optimize production processes, continually refining resource allocation, maintenance schedules, and infrastructure planning. Furthermore, these systems could simulate hypothetical scenarios, identifying untapped opportunities for clean energy generation and guiding strategic investments in emerging technologies, such as fusion reactors or advanced photovoltaic systems.

Regulatory Compliance and Sustainability Symbiosis: As environmental regulations evolve, generative AI could forge a symbiotic relationship with regulatory bodies. These systems could autonomously interpret and adapt to changing compliance requirements, generating tailored reports and recommendations for energy companies. Furthermore, GenAI will proactively identify potential regulatory shifts, empowering energy companies to stay ahead of the curve and shape sustainable practices that align with long-term environmental goals.

Generative AI presents a transformative opportunity for the energy industry's decarbonization efforts. By harnessing the power of recursive learning, predictive modeling, and autonomous optimization, these technologies could disrupt traditional approaches, unlocking new frontiers in sustainability, operational efficiency, materials science, and environmental stewardship.

Enabling a data strategy for AI

All these possibilities are highly dependent on data, both historical and synthetic data to train specialized FMs, but also real-time and current data to perform the inference driving predictions and automated decision making. Companies without modern data architectures fit for AI will struggle to realize the opportunities unleashed by generative AI and establishing a Data Store for LLMs is the first step. In the vast majority of cases, there is no need to copy the data, only to create what we call a knowledgebase (KB), which is a vector of your data in a vector database with a refresh routine to keep it up-to-date. Think about it as a real-time navigation app to where the data is located. Building KBs is direct and, depending on the complexity, can take a few days to a few weeks. These are not typically long development cycles.

There are four broad categories of data you will want to consider for your LLM data store:

1. **Data Platforms:** Systems such as OSDU and Data Warehouses designed to store large volumes of data.

2. **Conventional Enterprise Data:** Typically transaction-style applications that thrive on detailed transactions, just the kind of data an LLM needs.

3. **Third-Party Data Providers:** Paid data platforms from companies such as S&P, Enverus, and others who hold valuable data that your company licenses for a fee.

4. **Third-Party Data:** Public data sources that have no licensing fee and are generally available to the public.

The next step is to build agents that understand your data and KBs. Write a short script in Python that ingests the metadata about the dataset: what the data is used for, the schema of the data, the data structure and relationships, and how this KB might relate to other KBs. This is maybe 100 lines of code per agent. Bringing KBs and Agents together will provide a structure for your selected LLM to use your data without losing control of it.

Other key considerations

Having a clear business objective and a solid data foundation is critical, but they are not the only key considerations when it comes to generative AI. As generative AI matures and becomes widely adopted, companies will need to make sure of a responsible use of the technology, provide transparency of data lineage, guarantee security and privacy, and eliminate hallucinations to increase the autonomy and reliance on generative AI models.

FMs are expected to provide credible, reliable, and grounded responses that are based on public and proprietary source data. However, FMs can conflate multiple pieces of information, producing incorrect or new information, which impacts the reliability of the application. Reducing hallucinations needs to be an intentional area of focus and contextual grounding checks should be put in place. Hallucinations can be minimized with high quality training data, appropriate prompts, and model engineering. Technical providers are increasingly understanding the source of hallucinations and resolving them, especially in retrieval-augmented generation and conversational applications, and we expect they will eventually become rare events to manage.

Generative AI is the only available technology that can generate content, make decisions autonomously, and self-improve. As we have discussed, the value potential of the technology is immense, but so are the risks if it is used inappropriately. Yet balancing AI exploration with safety concerns is possible with collaboration between regulators and companies, and the adoption of safety standards. The risks of generative AI are quickly evolving and still emerging, and so the right technology, public-private collaboration, and adoption of controls need to be put in place ahead of large-scale deployments. Safeguards such as watermarking, white-hacker testing, Andon-cord mechanisms, understanding model limitations, raw data examination, adverse feedback modeling and testing, and input/output continuous monitoring need to be made sure of in each use case application.