Unlocking new insights with multimodal AI

Author: JOEY ZWICKER
General Manager,
AI Solutions, HPE

The next evolution of AI is here, and with it, bigger opportunities.

FROM REVOLUTIONIZING customer service to streamlining supply chains, AI has already transformed how we do business. But what we've seen so far is just the beginning. The next big leap in artificial intelligence is multimodal AI, and it promises even greater opportunities.

Multimodal AI refers to systems that can process and understand multiple types of data simultaneously. Unlike traditional unimodal AI, which typically focuses on a single type of data, multimodal AI integrates various data forms — such as text, images, and videos — to provide more comprehensive insights. For example, imagine a scenario where a multimodal AI system reads the phrase "a child making a shocked face" and simultaneously analyzes an image of a child with an expression of surprise. The AI understands that both inputs describe the same concept and can correlate and synthesize these different data types to gain deeper contextual understanding of the information.

Humans are inherently multimodal. We constantly process and interpret information from visual, textual, auditory, and other sources, often all at once. As we aim to develop AI models that can solve more complex and sophisticated problems, incorporating multimodal capabilities becomes essential. Multimodality enables AI systems to better mirror the multifaceted nature of human perception and understanding, driving a significant leap in AI sophistication.

Integrating diverse data types into a single AI model enhances its ability to make informed decisions, leading to better outcomes across various industries and use cases. In healthcare, for instance, a multimodal AI system can combine lab reports, doctor notes, MRI scans, and wearable data to provide a holistic view of a patient's health. This comprehensive approach allows for more accurate diagnoses and personalized treatment plans.

Another significant application is in government and defense, where threat analysis often involves processing text from social media, images from surveillance cameras, and videos from public sources. By synthesizing this information, multimodal AI can identify potential threats more accurately and efficiently than unimodal systems.

Overall, the ability to integrate and understand multiple data types is a key driver in advancing AI to tackle more complex and nuanced problems. This level of sophistication opens new avenues for innovation and decision-making effectiveness across various sectors, including finance, transportation, manufacturing, retail, and beyond.

Overcoming implementation challenges

Despite its potential, multimodal AI presents several challenges that need to be addressed. The complexity of processing and integrating multiple data types is significantly higher than handling a single data format. For instance, preprocessing images, text, and videos each require unique methods, and ensuring consistency and accuracy across these diverse datasets can be daunting.

Multimodality enables AI systems to better mirror the multifaceted nature of human perception and understanding, driving a significant leap in AI sophistication.

Additionally, the volume of data involved in multimodal AI applications is immense. Managing and processing such large datasets demands sophisticated infrastructure and advanced tools. Moreover, the untenability of humans manually labeling and verifying massive datasets introduces additional layers of complexity. This necessitates the development of highly efficient algorithms and preprocessing techniques to handle the scale and diversity of data effectively.

Another critical issue is the risk of bias. Multimodal AI models must be meticulously trained to avoid biases that can arise from the interplay of different data types. It's essential to ensure that the data is representative and that the model does not inadvertently learn and propagate incorrect behaviors or biases. This requires robust governance frameworks and data lineage tracking to maintain the integrity and reliability of the AI system.

Getting started with multimodal AI in your business

If you're interested in integrating multimodal AI into your operations, the first step is understanding where you stand in the AI adoption process. Assess whether your organization has the capabilities to train your own models or if it should leverage off-the-shelf models from trusted providers. Training and fine-tuning multimodal models is a complex, resource-intensive process that demands substantial investment in data preprocessing, model development, and validation.

If you opt to use existing models, be aware of the governance principles and risk factors associated with those models. Ensuring that the models meet your specific requirements and that you have robust mechanisms in place to manage potential biases and errors is crucial.

Another consideration is the strategic alignment of multimodal AI projects with your company's risk tolerance and ROI expectations. Multimodal AI can deliver transformative business value, but it also comes with significant risks and complexities. Engaging trusted advisors that have experience with high-scale AI systems can help you navigate these challenges and develop a realistic implementation roadmap.

Working with a trusted partner can also significantly stream-line the process of integrating multimodal AI into your business operations. Such partners bring valuable expertise and insights, helping you avoid common pitfalls and enabling a smoother implementation. They can provide comprehensive services that include architecture design, model training, and ongoing support, all tailored to your specific business needs.

Specifically, trusted advisors can assist in several critical areas:

- **TECHNOLOGY AND INFRASTRUCTURE:** They can help you choose the right tools and set up the necessary infrastructure to support large-scale data processing and AI model training.
- **DATA MANAGEMENT:** Experts can guide you in developing robust data management strategies, improving data quality, consistency, and security across all data streams.
- **GOVERNANCE AND COMPLIANCE:** Partners can assist in establishing governance frameworks to manage risks, maintain compliance, and promote the ethical use of AI.
- **CUSTOMIZATION AND INTEGRATION:** Advisors can tailor AI solutions to fit seamlessly with your existing systems and processes, maximizing the benefits of multimodal AI.

By leveraging the experience of these partners, you can reduce the time and cost associated with deploying multimodal AI and increase the likelihood of successful outcomes.

Harness the power of multimodal AI today

For business decision-makers, the time to start exploring and investing in multimodal AI is now. While the challenges are substantial, the potential rewards are even greater. By preparing your organization for this next wave of AI innovation, you can position your business to leverage multimodal AI for competitive advantage and drive future success.

About the author

Joey Zwicker is general manager for AI solutions at HPE. He co-founded Pachyderm, leading to its 2023 acquisition by HPE, and now drives AI software initiatives at HPE. He also chairs the AI Infrastructure Alliance, promoting AI innovation and standards for interoperability.

Hewlett Packard Enterprise

a50011354ENW