It is an annual report. Dear shareholders:  
As I sit down to write my second annual shareholder letter as CEO, I find myself optimistic and energized by what lies ahead for Amazon.

Chip development is a good example. In last year’s letter, I mentioned the investment we were making in our general-purpose CPU processors named Graviton. Graviton2-based compute instances deliver up to 40% better price-performance than the comparable latest generation x86-based instances; and in 2022, we delivered our Graviton3 chips, providing 25% better performance than the Graviton2 processors. Further, as machine learning adoption has continued to accelerate, customers have yearned for lower-cost GPUs (the chips most commonly used for machine learning). We delivered our first training chip in 2022 (“Trainium”); and for the most common machine learning models, Trainium-based instances are up to 140% faster than GPU-based instances at up to 70% lower cost. Most companies are still in the training stage, but as they develop models that graduate to  
large-scale production, they’ll find that most of the cost is in inference because models are trained periodically whereas inferences are happening all the time as their associated application is being exercised.  
One final investment area that I’ll mention, that’s core to setting Amazon up to invent in every area of our business for many decades to come, and where we’re investing heavily is Large Language Models (“LLMs”) and Generative AI. This shift was driven by several factors, including access to higher volumes of compute capacity at lower prices than was ever available. Amazon has been using machine learning extensively for 25 years, employing it in everything from personalized ecommerce recommendations, to fulfillment center pick paths, to drones for Prime Air, to Alexa, to the many machine learning services AWS offers (where AWS has the broadest machine learning functionality and customer base of any cloud provider).



