



Werner Vogels' Tech Predictions for 2024 and Beyond

Dr. Werner Vogels
VP & CTO, Amazon.com

Contents

- Overview 3
- Generative AI becomes culturally aware 4
- FemTech finally takes off 7
- AI assistants redefine developer productivity10
- Education evolves to match the speed of technology13

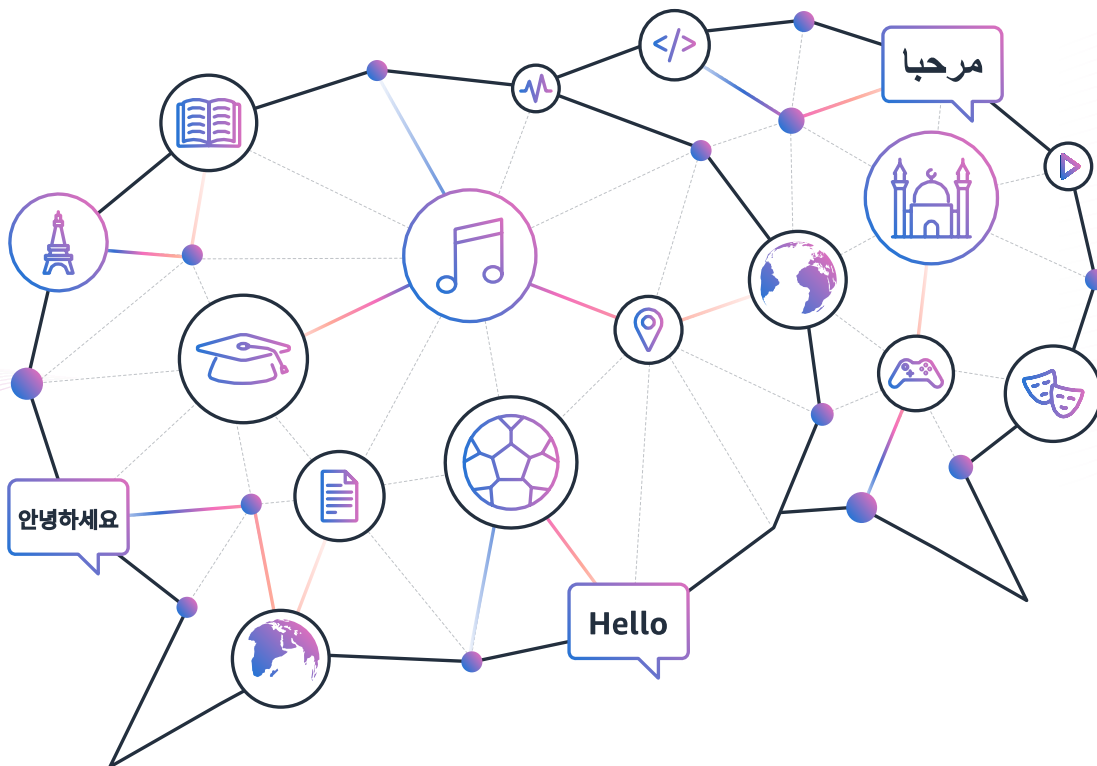


Overview

Throughout history, people have developed tools and systems to augment and amplify their own capabilities. Whether the printing press or the assembly line, these innovations have allowed us to do more than we ever could alone. Jobs changed, new professions emerged, and people adapted. In the past year, the rate of change has rapidly accelerated. Cloud technologies, machine learning, and generative AI have become more accessible, impacting nearly every aspect of our lives from writing emails to developing software, even detecting cancer at an early stage. The coming years will be filled with innovation in areas designed to democratize access to technology and help us keep up with the increasing pace of every-day life—and it starts with Generative AI.

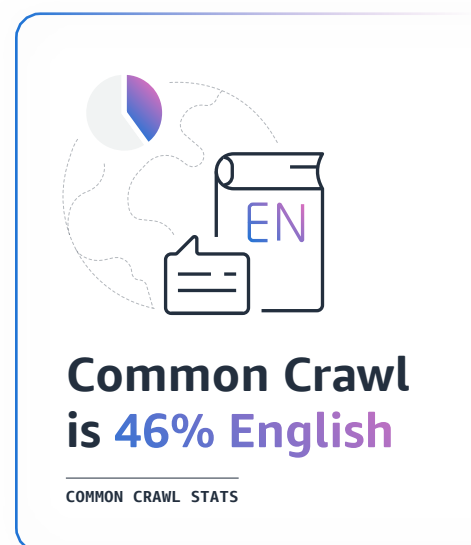
Generative AI becomes culturally aware

Large language models (LLMs) trained on culturally diverse data will gain a more nuanced understanding of human experience and complex societal challenges. This cultural fluency promises to make generative AI more accessible to users worldwide.



Culture influences everything. The stories we tell, the food we eat, the way we dress, our values, our manners, our biases, the way we approach problems and make decisions. It is the foundation for how each one of us exists within a community. Culture provides rules and guidelines that inform and govern our behaviors and beliefs—and this contract changes depending on where we are and who we are with. At the same time, these differences can sometimes result in confusion and misinterpretation. In Japan, it is considered a sign of enjoyment to slurp your soup as you eat noodles, but it is considered impolite in other cultures. At a traditional wedding in India, a bride may wear an intricately designed and colorful lehenga, while in the western world the tradition is a white dress. And in Greece it is customary to spit on the dress for good luck. As humans, we are used to working across cultures, and as such, we can contextualize this information, synthesize it, adjust our understanding and respond appropriately. So, why would we expect anything less from the technologies that we use and rely on in our daily lives? In the coming years, culture will play a crucial role in how technologies are designed, deployed, and consumed; its effects will be most evident in generative AI.

For LLM-based systems to reach a world-wide audience, they need to achieve the type of cultural fluency that comes instinctively to humans. In a paper published earlier this year, [researchers from Georgia Institute of Technology demonstrated](#) that even when an LLM was provided with a prompt in Arabic that explicitly mentioned Islamic prayer, responses were generated that recommended grabbing an alcoholic beverage with friends, which isn't culturally appropriate. A lot of this has to do with the training data that's available. Common Crawl, which has been used to train many LLMs, is [roughly 46% English](#), and an even greater percentage of the content available—regardless of language—is culturally Western (skewing significantly towards the United States). Using the same prompt with a [model pre-trained on Arabic texts, specifically for Arabic language generation](#), culturally appropriate responses were generated, such as grabbing a tea or coffee. In the past few months, non-Western LLMs have started to emerge: Jais, trained on Arabic and English data, Yi-34B, a bilingual Chinese/English model, and Japanese-large-lm, trained on an extensive Japanese web corpus. These are signs that culturally accurate non-Western models will open up generative AI to hundreds of millions of people with impacts ranging far and wide, from education to medical care.

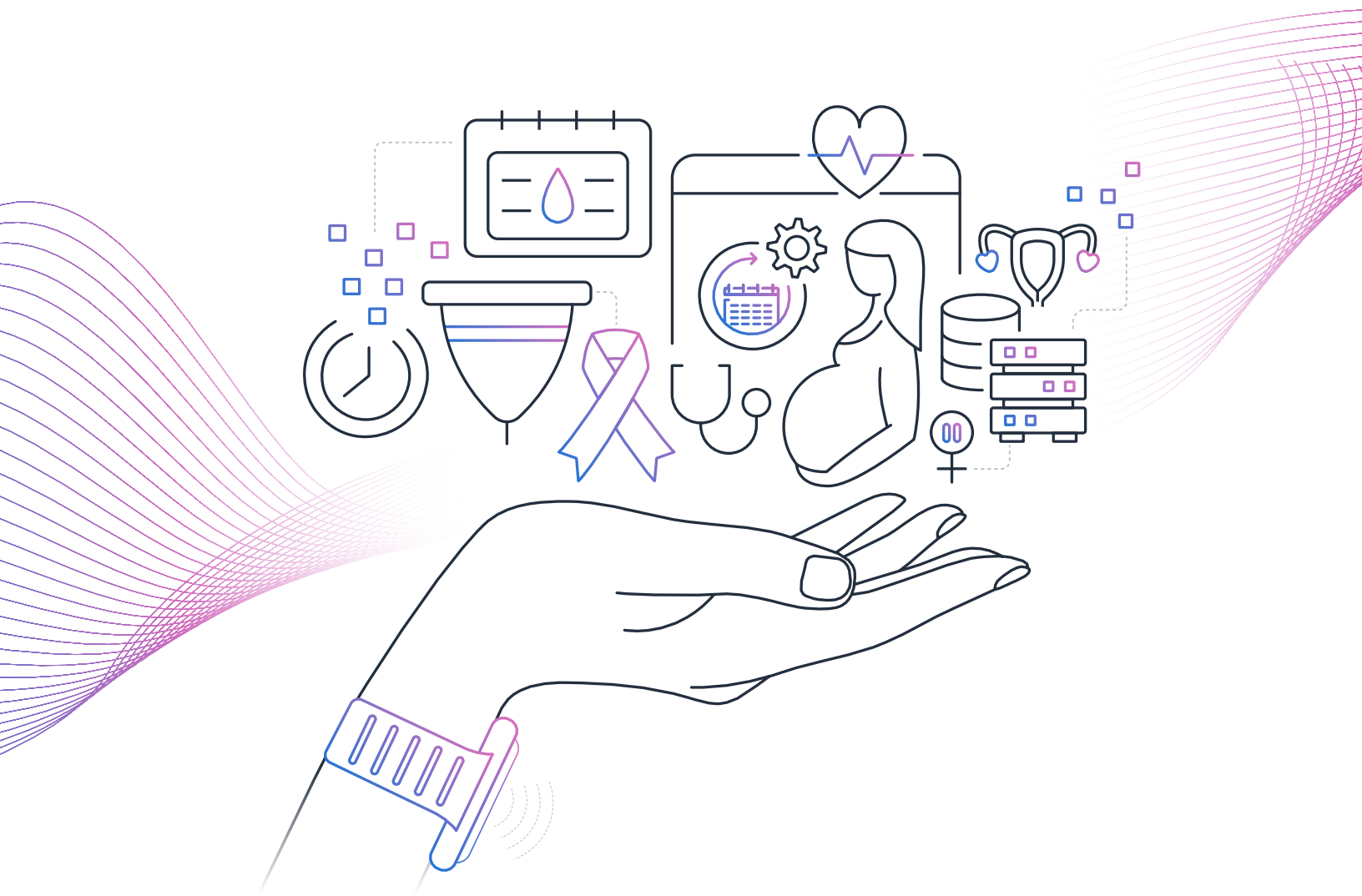


Keep in mind, language and culture are not the same. Even being able to do perfect translation does not give a model cultural awareness. As a myriad of histories and experiences are embedded into these models, we will see LLMs begin to develop a broader, worldwide range of perspectives. Just as humans learn from discussion, debate, and the exchange of ideas, LLMs need similar opportunities to expand their perspectives and understand culture. Two areas of research will play a pivotal role in this cultural exchange. One is reinforcement learning from AI feedback (RLAIF), in which a model incorporates feedback from another model. In this scenario, different models can interact with each other and update their own understandings of different cultural concepts based on these interactions. Second is collaboration through multi-agent debate, in which multiple instances of a model generate responses, debate the validity of each response and the reasoning behind it, and finally come to an agreed upon answer through this debate process. Both areas of research reduce the human cost it takes to train and fine-tune models.

As LLMs interact and learn from each other, they will gain more nuanced understandings of complex societal challenges informed by diverse cultural lenses. These advances will also ensure that models provide more robust and technically accurate responses across a broad range of topics in areas such as tech. The effects will be profound and felt across geographic regions, communities, and generations to come.

FemTech finally takes off

Women's healthcare reaches an inflection point as FemTech investment surges, care goes hybrid, and an abundance of data unlocks improved diagnoses and patient outcomes. The rise of FemTech will not only benefit women, but lift the entire healthcare system.



Women's healthcare is not a niche market. In the United States alone, women spend more than \$500 billion per year on care. They make up 50% of the population and account for 80% of consumer healthcare decisions. However, the foundation of modern medicine has been male by default. It wasn't until the [NIH Revitalization Act of 1993](#) that women in the US were even included in clinical research. Common needs like menstrual care and menopause treatment have historically been treated as taboo, and because women have been excluded from trials and research their outcomes have typically been worse than men. On average, women are diagnosed later than men for many diseases, and women are [50%](#) more likely to be misdiagnosed following a heart attack. Maybe the most glaring example of the inequities is prescription medicine, where women report adverse side effects at significantly higher rates than men. Though these statistics seem concerning on the surface, investment in women's healthcare (aka FemTech) is on the rise, aided by cloud technologies and greater access to data.

At AWS, we've been working closely with women-led start-ups and have seen first-hand the growth in FemTech. In the last year alone, funding has increased 197%. With increased access to capital, technologies like machine learning, and connected devices designed specifically for women, we are at the precipice of an unprecedented shift, not only in the way women's care is perceived, but how it's administered. Companies like [Tia](#), [Elvie](#), and [Embr Labs](#) are showing the immense potential of leveraging data and predictive analytics to provide individualized care and meet patients where they're comfortable—at home and on-the-go.

As stigma fades around women's health needs and more funding flows into the sector, we will see FemTech companies continue to aggressively tackle previously overlooked conditions and needs. At the same time, women's access to health services will dramatically increase thanks to hybrid care models that take advantage of online medical platforms, the availability of low-cost diagnostic devices, and on-demand access to medical professionals. Customers like [Maven](#) have proven themselves to be leaders in this space, blurring the lines between mental health and physical wellbeing, providing everything from relationship counseling to menopause care. As these platforms mature and proliferate, we will see access to care democratized. Women in rural areas and

historically underserved regions will have an easier time connecting to OB/GYNs, mental health professionals, and other specialists through apps and telehealth platforms. Smart tampon systems like the one [NextGen Jane](#) is developing, will let women establish profiles of their uterine health and identify potential genomic markers of disease, which can be seamlessly shared with their clinicians. And wearables will provide users and their doctors with an abundance of longitudinal health data that can be analyzed. Where today, more than 70% of women go untreated for menopause symptoms, increased education, availability of data, and non-invasive solutions will dramatically improve outcomes—and it goes well beyond OB/GYN care.

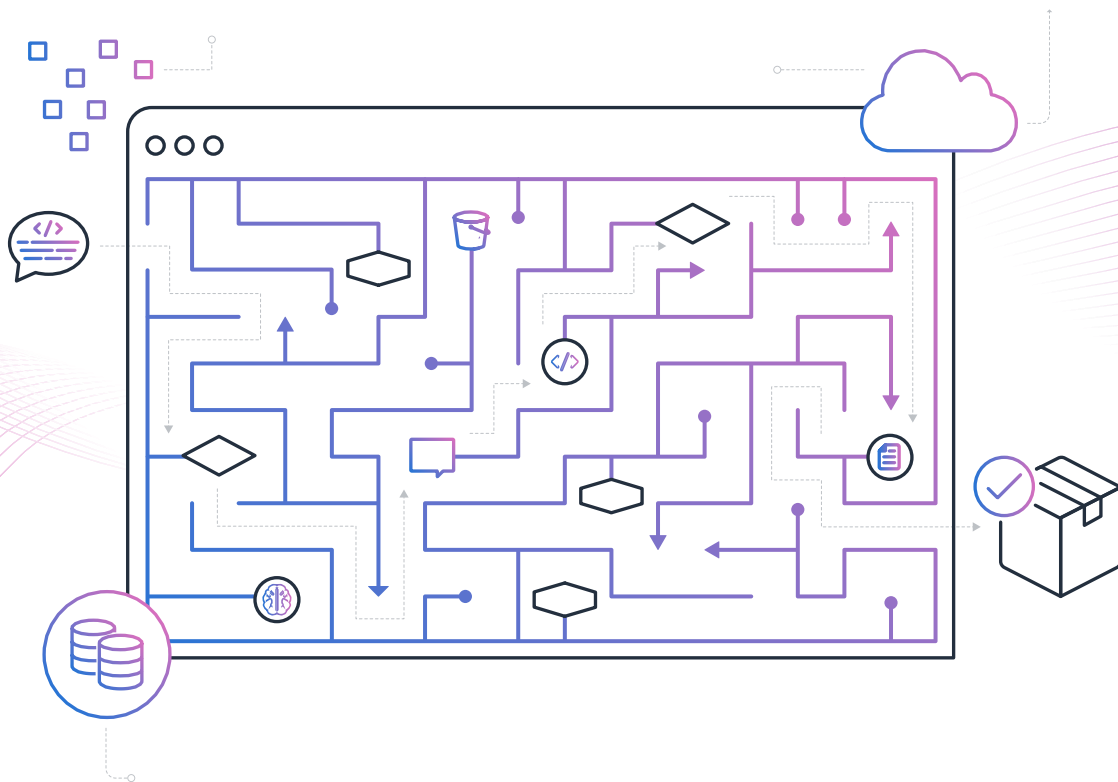


For example, in the run-up to the Women’s World Cup, [roughly 30 athletes suffered tournament ending ACL injuries](#). Like with traditional medicine, women’s training was modeled on what worked for men without much consideration for physiology. As a result, women have been six times as likely to go down with an ACL injury and 25% less likely to make a full recovery and return to the pitch. This is another area where understanding unique health data will have an impact, not only to prevent injuries, but to improve the health of women athletes holistically.

We are at an inflection point for women’s healthcare. Access to an abundance of diverse data coupled with cloud technologies, like computer vision and deep learning, will reduce misdiagnoses and help minimize medication side effects that disproportionately impact women today. Endometriosis and postpartum depression will receive the attention they rightfully deserve. We’ll finally see women’s care move from the fringes to the forefront. And [since women-led teams are more inclined than those made up of just men to solve a broad range of health issues](#), we’ll see FemTech not only benefit those who identify as women, but lift the entire healthcare system.

AI assistants redefine developer productivity

AI assistants will evolve from basic code generators into teachers and tireless collaborators that provide support throughout the software development lifecycle. They will explain complex systems in simple language, suggest targeted improvements, and take on repetitive tasks, allowing developers to focus on the parts of their work that have the most impact.



In 2021, I predicted that generative AI would start to play a major role in the way software was written. It would augment the developers' skills, helping them write more secure and reliable code. We are seeing exactly that in earnest now, with broad access to tools and systems that can generate entire functions, classes, and tests based on natural language prompts. In fact, in the [2023 Stack Overflow Developer Survey](#), 70% of respondents said they were already using or planning to use AI supported tools in their development processes.



The AI assistants on the horizon will not only understand and write code, they will be tireless collaborators and teachers. No task will exhaust their energy, and they'll never grow impatient explaining a concept or redoing work—no matter how many times you ask. With infinite time and unlimited patience, they will support everyone on the team and contribute to everything from code reviews to product strategy.

The lines between product managers, front- and back-end engineers, DBAs, UI/UX designers, DevOps engineers, and architects will blur. With contextual understanding of entire systems, not just isolated modules, AI assistants will provide recommendations that augment human creativity, such as translating a napkin sketch into scaffolding code, generating templates from a requirements doc, or recommending the best infrastructure for your task (e.g., serverless vs. containers).

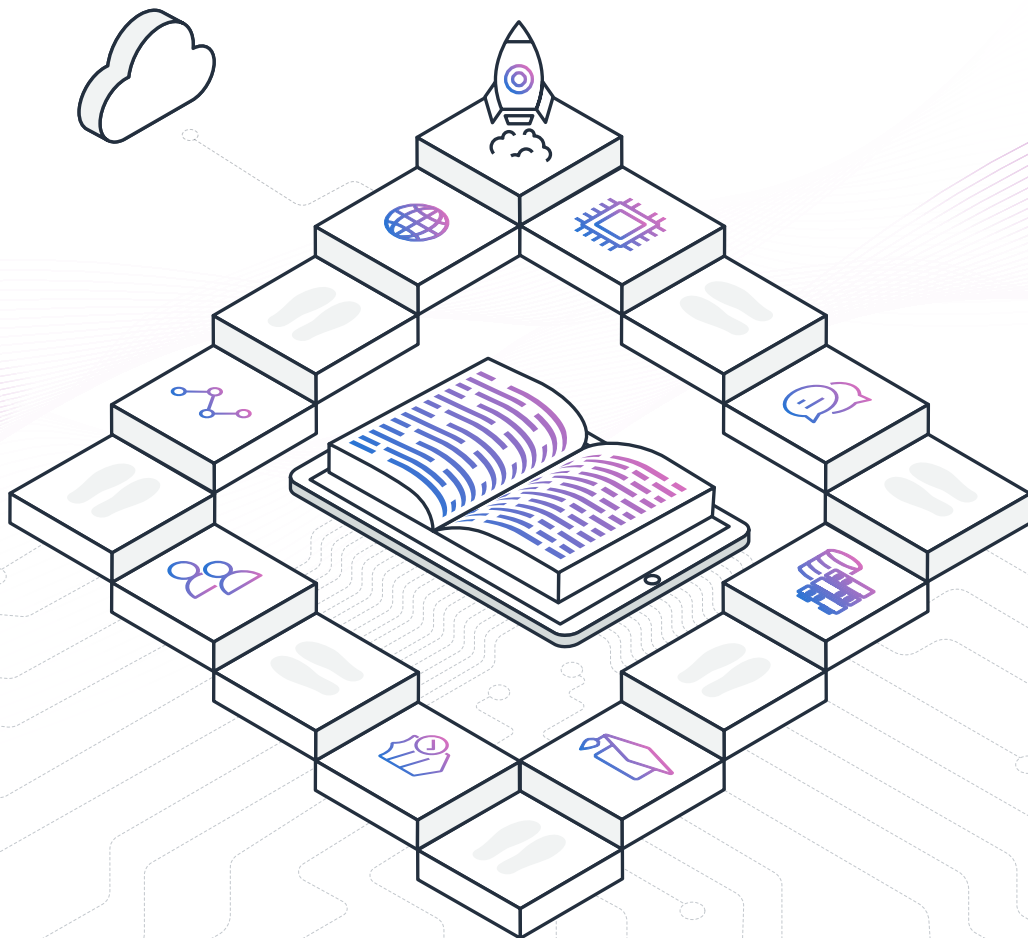
These assistants will be highly customizable—personalized at the individual, team, or company level. They'll be able to explain the internals of complex distributed systems, like Amazon S3, in simple terms, making them invaluable educational tools. Junior developers will leverage them to quickly get up to speed on unfamiliar infrastructure. Senior engineers will use them to swiftly comprehend new projects or codebases and begin making meaningful contributions. Whereas before it may have taken weeks to fully grasp the downstream impacts of a code change, assistants can instantly assess modifications, summarize their effects on other parts of the system, and suggest additional changes as needed.

We are already seeing some of the most tedious parts of modern software development taken off the plates of developers: writing unit tests, boilerplate code, and debugging errors. The tasks that are often considered “extra” and fall by the wayside. These assistants will be able to re-architect and migrate entire legacy applications, such as upgrading from Java 8 to 17, or decomposing from a monolith into microservices. Make no mistake, developers will still need to plan and evaluate outputs. But these assistants will help sift through academic research and choose the right algorithm for your distributed system, determine how to best move from a primary-backup approach to an active-active implementation, even understand how resources individually impact efficiency and develop pricing models. As a result, there will be more work than ever. Unburdened by the undifferentiated heavy lifting of tasks like upgrading Java versions, developers can focus on the creative work that drives innovation.

In the coming years, engineering teams will become more productive, develop higher quality systems, and shorten software release lifecycles as AI assistants move from novelty to necessity across the entire software industry.

Education evolves to match the speed of technology

Higher education alone cannot keep up with the rate of technological change. Industry-led skills-based training programs will emerge that more closely resemble the journeys of skilled tradespeople. This shift to continuous learning will benefit individuals and businesses alike.



I remember the software development cycles of the past, when a product might be in development for 5+ years before ever reaching a customer's hands. In the late-90s, this was an acceptable approach. But in today's world, this software would be severely outdated before ever being put to any real use. Because of access to cloud computing, a culture of continuous improvement, and the widespread adoption of the minimum viable product approach, our software development cycles have shortened. And the impact has been significant. Companies are bringing products to market faster than ever and customers are adopting new technologies at previously unimaginable speeds. In this rapidly spinning flywheel of technology and business, one area that has not been included until now, is higher education.

Education is radically different across the world, but it's been widely accepted that to hire the best people—and to land the best job yourself—a college degree is table stakes. This has been especially true in technology. But we're beginning to see this model break down, both for individuals and for companies. For students, costs are rising and many are questioning the value of a traditional college degree when practical training is available. For companies, fresh hires still require on-the-job-training. As more and more industries call for specialization from their employees, the gap is widening between what's taught in school and what employers need. Similar to the software development processes of decades past, we have reached a pivotal point with tech education, and we will see what was once bespoke on-the-job-training for a few evolve into industry-led skills-based education for many.

We have seen glimpses of this shift underway for years. Companies like Coursera, who originally focused on consumers, have partnered with enterprises to scale their upskilling and reskilling efforts. Degree apprenticeships have continued to grow in popularity, because education can be specialized by the employer, and apprentices can earn as they learn. But now, companies themselves are starting to seriously invest in skills-based education at scale. In fact, Amazon just announced that it has already trained **21 million learners** across the world in tech skills. And it's in part thanks to programs like the [Mechatronics and Robotics Apprenticeship](#) and [AWS Cloud Institute](#). All of these programs enable learners at different points in their career journey to gain the exact skills they need to enter in-demand roles, without the commitment of a traditional multi-year program.



21 million tech learners trained worldwide by Amazon

ABOUT AMAZON

To be clear, this concept is not without precedent: when you think about skilled workers like electricians, welders, and carpenters, the bulk of their skills are not gained in the classroom. They move from trainee to apprentice to journeyman, and possibly master tradesperson. Learning is continuous on the job, and there are well defined paths to upskill. This style of lifelong education—to learn and be curious—bodes well for individuals and businesses alike.

None of this means that traditional degrees are going away. This is not an “either/or” situation—it’s about choice. There will still be areas in tech where this type of academic learning is critical. But there will be many industries where the impact of technology outpaces traditional educational systems. To meet the demands of business, we will see a new era of industry-led educational opportunities that can’t be ignored.

