# Splash Music accelerates AI-driven music creation with AWS Trainium and HyperPod



In the rapidly evolving world of artificial intelligence, Splash Music is emerging as a trailblazer by leveraging advanced cloud infrastructure to transform music creation. The startup’s HummingLM model, designed to convert hummed inputs into fully realised musical tracks, exemplifies how AI is redefining creative workflows beyond traditional algorithmic music generation. This innovation is closely tied to Splash Music’s collaboration with Amazon Web Services (AWS), particularly through its involvement in the 2024 AWS Generative AI Accelerator program, which has accelerated the model’s development and deployment capabilities.

Central to Splash Music’s breakthrough is the utilisation of AWS Trainium chips and Amazon SageMaker HyperPod infrastructure. Trainium is AWS’s family of AI chips engineered to deliver superior training performance at significantly reduced costs. These chips empower the complex, multi-stage processes within HummingLM, such as encoding audio hums and expanding them into intricate musical compositions with rhythm, melody, and instrumentation. Industry data show that Trainium-powered setups can cut training costs by up to 50% compared to other options, enabling startups like Splash to train larger models efficiently and economically. Additionally, SageMaker HyperPod facilitates scalable, resilient cluster management with automated failure recovery, ensuring near-zero downtime—a critical factor for a startup relying on rapid iterative cycles to stay competitive. This integration has reportedly reduced Splash’s operational training costs by over 54%, allowing frequent model updates on a weekly basis and fast innovation in generative music services.

The underlying technological sophistication of HummingLM mirrors advances seen in other AI music generation projects such as Google’s MusicLM and Meta’s MusicGen. However, Splash Music distinguishes itself with a user-focused approach that emphasises real-time music creation from simple hummed inputs—a capability underscored by the efficiency of their AWS-powered training environment. Notably, the combination of Trainium chips and SageMaker HyperPod has reduced generation times substantially, with improvements in cluster resilience and resource management enabling seamless scalability across diverse musical genres and styles. These infrastructure efficiencies correspond with broader industry trends highlighted in recent AWS announcements, including the launch of Trainium3 chips and AI tools like Amazon Nova, all designed to enhance generative AI performance and scalability.

The industry implications of Splash Music’s approach could be profound. By democratising access to sophisticated AI music production tools, smaller artists and independent creators can potentially rival traditional studios in quality and innovation. This disruption is aided by the significantly lower costs and rapid iteration enabled by AWS hardware and software advances. Moreover, the ongoing enhancements in generative AI infrastructure signal a future where AI-assisted creativity extends beyond music into multimodal domains, blending voice, text, and visual inputs to craft richer artistic outputs. While challenges around ethical AI use and copyright in generated music remain crucial considerations, Splash Music’s model exemplifies how well-supported AI innovation can blend human creativity with machine precision effectively.

Competition in the AI music generation space is intensifying, with projects like DeepMind’s Magenta Real-time pushing impressive benchmarks, including handling large-scale audio datasets through massive transformer models. Yet Splash Music’s humming-first approach, optimised for real-time interaction and supported by distributed training on HyperPod, offers a compelling alternative tailored for interactive user experiences. The strategic partnership with AWS not only provides cost-efficient scalability but also aligns with broader cloud industry movements to integrate domain-specific foundation models with powerful computing infrastructure. This synergy between cloud technology and AI creativity underscores a broader trend where targeted infrastructure investments accelerate niche innovations, fostering an expansive new toolkit for artists and developers alike.

Ultimately, Splash Music’s journey illustrates how the convergence of specialised AI hardware, sophisticated software frameworks, and collaborative ecosystem support can unlock new frontiers in digital creativity. As AWS continues to enhance its AI offerings—further optimised by Trainium3 chips and advanced distributed training systems—startups like Splash are well positioned to lead the next wave of creative technology. This emerging landscape promises to reshape industries by merging human artistic vision with the precision and scalability of generative AI, heralding a new era in how music and other art forms are conceived and produced.

### 📌 Reference Map:

* Paragraph 1 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[2]](https://aws.amazon.com/blogs/machine-learning/splash-music-transforms-music-generation-using-aws-trainium-and-amazon-sagemaker-hyperpod/)
* Paragraph 2 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[2]](https://aws.amazon.com/blogs/machine-learning/splash-music-transforms-music-generation-using-aws-trainium-and-amazon-sagemaker-hyperpod/), [[4]](https://aws.amazon.com/machine-learning/trainium//), [[5]](https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod.html)
* Paragraph 3 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[4]](https://aws.amazon.com/machine-learning/trainium//), [[5]](https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod.html), [[6]](https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod-recipes-tutorials.html), [[3]](https://www.aboutamazon.com/news/aws/amazon-nova-ai-canvas-reel-aws-reinvent)
* Paragraph 4 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[2]](https://aws.amazon.com/blogs/machine-learning/splash-music-transforms-music-generation-using-aws-trainium-and-amazon-sagemaker-hyperpod/), [[3]](https://www.aboutamazon.com/news/aws/amazon-nova-ai-canvas-reel-aws-reinvent)
* Paragraph 5 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[7]](https://arxiv.org/abs/2404.10630)
* Paragraph 6 – [[1]](https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/), [[2]](https://aws.amazon.com/blogs/machine-learning/splash-music-transforms-music-generation-using-aws-trainium-and-amazon-sagemaker-hyperpod/), [[4]](https://aws.amazon.com/machine-learning/trainium//), [[5]](https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod.html), [[3]](https://www.aboutamazon.com/news/aws/amazon-nova-ai-canvas-reel-aws-reinvent)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.webpronews.com/splash-musics-humminglm-turns-hums-into-full-tracks-with-aws-ai/> - Please view link - unable to able to access data
2. <https://aws.amazon.com/blogs/machine-learning/splash-music-transforms-music-generation-using-aws-trainium-and-amazon-sagemaker-hyperpod/> - Splash Music has revolutionised music generation by deploying its HummingLM model on AWS Trainium chips and Amazon SageMaker HyperPod. This collaboration has led to automated, resilient, and scalable training, reducing operational downtime to near zero and enabling weekly model refreshes. The integration has also resulted in over a 54% reduction in training costs, allowing Splash Music to train larger models and accelerate innovation in generative music services.
3. <https://www.aboutamazon.com/news/aws/amazon-nova-ai-canvas-reel-aws-reinvent> - At AWS re:Invent 2024, Amazon introduced several advancements, including Amazon Nova, Amazon Bedrock innovations, and Trainium3 chips. These developments aim to enhance the performance and scalability of generative AI workloads, providing customers with more powerful tools for AI model development and deployment.
4. <https://aws.amazon.com/machine-learning/trainium//> - AWS Trainium is a family of AI chips purpose-built by AWS for AI training and inference, delivering high performance while reducing costs. The first-generation AWS Trainium chip powers Amazon EC2 Trn1 instances, which have up to 50% lower training costs than comparable Amazon EC2 instances. AWS Trainium2 chip delivers up to 4x the performance of first-generation Trainium, and Trainium2-based Amazon EC2 Trn2 instances are purpose-built for generative AI and are the most powerful EC2 instances for training and deploying models with hundreds of billions to trillion+ parameters.
5. <https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod.html> - Amazon SageMaker HyperPod helps provision resilient clusters for running machine learning workloads and developing state-of-the-art models such as large language models (LLMs), diffusion models, and foundation models (FMs). It accelerates development of FMs by removing undifferentiated heavy-lifting involved in building and maintaining large-scale compute clusters powered by thousands of accelerators such as AWS Trainium and NVIDIA A100 and H100 GPUs.
6. <https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-hyperpod-recipes-tutorials.html> - The tutorials provided by Amazon SageMaker AI offer quick-start guides to using the recipes for training, including pre-training and fine-tuning with SageMaker HyperPod. These tutorials help users get started with training and fine-tuning publicly available foundation models (FMs) in minutes with state-of-the-art performance.
7. <https://arxiv.org/abs/2404.10630> - The paper titled 'HLAT: High-quality Large Language Model Pre-trained on AWS Trainium' showcases a family of 7B and 70B decoder-only LLMs pre-trained using 4096 AWS Trainium accelerators over 1.8 trillion tokens. The performance of HLAT is benchmarked against popular open-source models, demonstrating that AWS Trainium, powered by NeuronX Distributed Training (NxDT), can successfully pre-train state-of-the-art LLM models with high performance and cost-effectiveness.