Trade Journal Entry #2

TechCrunch, AI, Sept. 26th to Oct. 3rd

In TechCrunch AI section, over 80 articles are published during last week including following (only shows seven about different topics):

* Governor Newsom announces new initiatives to advance safe and responsible AI, protect Californians
* Here is what’s illegal under California’s 18 (and counting) new AI laws
* Y Combinator is being criticized after it backed an AI startup that admits it basically cloned another AI startup
* Microsoft Copilot can now read your screen, think deeply, and speak aloud to you
* Anthropic hires OpenAI co-founder Durk Kingma
* Meta won’t say whether it trains AI on smart glasses photos
* Raspberry Pi launches camera module for vision-based AI applications

This week in the AI section of TechCrunch, several notable developments took place. California Governor Gavin Newsom vetoed SB 1047, continuing to position California at the forefront of AI legislation. However, other AI-related regulations, such as those addressing deepfake nudes and AI training data disclosures, were signed into law. Meanwhile, Microsoft upgraded Copilot, which can now "read screens, think deeply, and even speak aloud to users." Another significant industry shift saw Durk Kingma, one of OpenAI's co-founders, joining Anthropic. In the AI hardware space, Raspberry Pi released its AI Camera product with onboard AI processing, while startup Cerebras made headlines by filing for an IPO despite reporting heavy losses.

A major event this week was OpenAI’s 2024 Dev Day. ( <https://techcrunch.com/2024/10/01/openais-devday-brings-realtime-api-and-other-treats-for-ai-app-developers/> ) During the event, OpenAI introduced several tools aimed at expanding developer capabilities. One key highlight was the launch of the Realtime API, which enables low-latency, AI-generated voice responses for real-time speech-to-speech interactions—ideal for applications like trip planners or food ordering services. Another major release was vision fine-tuning, allowing developers to personalize and enhance pre-trained GPT-4o models using images, improving AI performance in visual tasks while applying filters to restrict copyrighted or violent content. OpenAI also introduced model distillation, allowing developers to fine-tune smaller models using larger ones like GPT-4o. This offers both cost-efficiency and performance improvements. Additionally, prompt caching was introduced, reducing costs and improving latency by storing frequently used context between API calls. The introduction of model distillation provides companies with fewer resources an opportunity to integrate AI by making smaller, cost-effective models perform almost as well as larger ones.

These updates from OpenAI indicate strong potential for the company’s future development. The Realtime API and vision fine-tuning open new possibilities for developers to create more interactive, responsive applications, especially in areas like customer service, travel, and AI assistants. These features make it easier to build complex apps with lower latency and better visual understanding, addressing key demands from developers. However, the absence of updates on OpenAI’s video generation model, Sora, or the GPT Store could allow competitors like Anthropic or Meta to gain ground. By focusing on refining existing tools rather than introducing entirely new models, OpenAI seems to be shifting toward practical, developer-centric enhancements, signaling that the broader AI industry may enter a phase of consolidation—focused on optimizing and scaling AI for everyday use, rather than chasing flashy, groundbreaking innovations.