Four AI trends to watch

March 21 2023, by Rishi Bommasani

Credit: Pixabay/CC0 Public Domain

This was a major week in AI, with some of the largest players releasing flagship models: OpenAI's highly anticipated GPT-4 made its debut, while Google put out its Med-PaLM 2, a new-and-improved medical model, as well as their PaLM API and MakerSuite to enable developers to prototype and deploy their own models. Meanwhile, Anthropic

introduced its ChatGPT competitor, Claude API, in an invite-only release of its chatbot and a lighter, cheaper version (Claude Instant). And in China, Tsinghua launched ChatGLM, a chat-based Chinese-English model reminiscent of ChatGPT.

In this AI spring, much awaits. Technology rapidly improves, leading to unprecedented adoption and more funding. But transparency dwindles as companies grow more protective of their products. Amidst all this activity, policy interest blooms, but concrete action lags behind.

Trend 1: Deployment on the rise

November's release of ChatGPT turned AI mainstream. Two months after launch, it amassed 100 million monthly active users, making it the fastest-growing consumer application in history. Companies quickly incorporated foundation models into their public-facing products: Earlier this month, OpenAI released the ChatGPT API, highlighting how Snap and Quizlet already ship products based on ChatGPT and Instacart announcing plans for later this year. OpenAI and Salesforce partnered to incorporate ChatGPT into the workplace software Slack.

In the past week, the deployment has further accelerated: Anthropic's Claude now powers Notion AI and the search engine DuckDuckGo, while OpenAI's GPT-4 underpins offerings at Morgan Stanley, Khan Academy, Duolingo, and Stripe. In fact, OpenAI's release details ongoing efforts with the government of Iceland for language preservation. Meanwhile, the online Q&A forum Quora released Poe, a chatbot service that offers both Anthropic and OpenAI models on the backend.

Finally, both Google and Microsoft put out plans to sweepingly deploy foundation models in many of their iconic products from Google Slides to Microsoft Word.

Trend 2: Worsening transparency

As AI grows more capable and useful, many leading organizations have turned less transparent.

Nothing typifies this more than OpenAI's technical report on GPT-4 that reads, "Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar." Here at Stanford's Center for Research on Foundation Models (CRFM), we have argued for the urgent need for robust norms.

However, we do see clear efforts to combat these trends, highlighting several transparent releases. Meta publicly released LLaMA with license requirements on commercial use. Together unveiled OpenChatKit 0.15, a chatbot that developers can build from, with full access to source code, model weights, and training datasets. Tsinghua's ChatGLM-6B is also open-source.

At CRFM, we released the training recipe for Alpaca, based on Meta's LLaMa-7B that significantly reduces the cost to acquire capabilities comparable to OpenAI's GPT-3.5 series models (text-davinci-003 in particular).

Trend 3: Massive influx of funding

Underpinning all this recent attention is a massive torrent of funding on the back of Microsoft's $10 billion influx to OpenAI. The past few months have been dubbed a "gold rush" for foundation model startups, capitalizing on the interest in generative AI and sky-high valuations. Some standouts: Adept raised $350 million, Google dropped another

$300 million into Anthropic, Character AI raised $200 million from Andreesen Horowitz, Perplexity raised $25 million, and Salesforce Ventures launched a $250 million fund, supporting a bevy of startups including Anthropic, Cohere, and You.com. While recent funding attention has centered on text, what will we see in the coming year across other modalities (e.g. Stability (the developers of StableDiffusion) raised $101 million in October)?

Trend 4: Demand for policy

As foundation models go mainstream, the demand for AI policy grows, capturing the attention of top policymakers across the world. In the U.S., Congressman Ted Lieu proposed a resolution calling for AI regulation that was written by ChatGPT, accompanied by an op-ed in the *New York Times*. While governments have been slow to pass legislation regulating artificial intelligence, traction is picking up. Keying in on the poor transparency, CRFM recently put forth a policy brief calling for standardized evaluations and public reporting through efforts like HELM .

In late 2022, the White House shared a blueprint for an AI bill of rights, and last month a White House-led task force behind the National AI Research Resource (NAIRR) issued its final report, asking Congress for a six-year investment to build out AI resources. MIT Professor Aleksander Madry testified to the U.S. House Subcommittee on Cybersecurity, Information Technology, and Government Innovation last week: "AI is no longer a matter of science fiction, nor is it a technology confined to research labs. AI is a technology that is already being deployed and broadly adopted as we speak. It will drastically change our lives; we need to be thinking now how to shape the AI-driven world that comes."

In the past week, the U.K. announced plans to launch a task force on

foundation models, reporting directly to the prime minister, followed already by £900 million investment into an exascale supercomputer. As Prime Minister Rishi Sunak put it, "Foundation models like ChatGPT are beginning to demonstrate remarkable new abilities. … It is our new reality. And the race to create, develop, and exploit these new technologies is global."

Provided by Stanford University