



Generative AI: A Creative New World

By [Sonya Huang](#), [Pat Grady](#) and GPT-3

Published September 19, 2022

A powerful new class of large language models is making it possible for machines to write, code, draw and create with credible and sometimes superhuman results.

Humans are good at analyzing things. Machines are even better. Machines can analyze a set of data and find patterns in it for a multitude of use cases, whether it's fraud or spam detection, forecasting the ETA of your delivery or predicting which TikTok video to show you next. They are getting smarter at these tasks. This is called "Analytical AI," or traditional AI.

But humans are not only good at analyzing things—we are also good at creating. We write poetry, design products, make games and crank out code. Up until recently, machines had no chance of competing with humans at creative work—they were relegated to analysis and rote cognitive labor. But machines are just starting to get good at creating sensible and beautiful things. This new category is called "Generative AI," meaning the machine is generating something new rather than analyzing something that already exists.

Generative AI is well on the way to becoming not just faster and cheaper, but better in some cases than what humans create by hand. Every industry that requires humans to create original work—from social media to gaming, advertising to architecture, coding to graphic design, product design to law, marketing to sales—is up for reinvention. Certain functions may be completely replaced by generative AI, while others are more likely to thrive from a tight iterative creative cycle between human and machine—but generative AI should unlock better, faster and cheaper creation across a wide range of end markets. The dream is that generative AI brings the marginal cost of creation and knowledge work down towards zero, generating vast labor productivity and economic value—and commensurate market cap.

The fields that generative AI addresses—knowledge work and creative work—comprise billions of workers. Generative AI can make these workers at least 10% more efficient and/or creative: they become not only faster and more efficient, but more capable than before. Therefore, Generative AI has the potential to generate trillions of dollars of economic value.

Why Now?

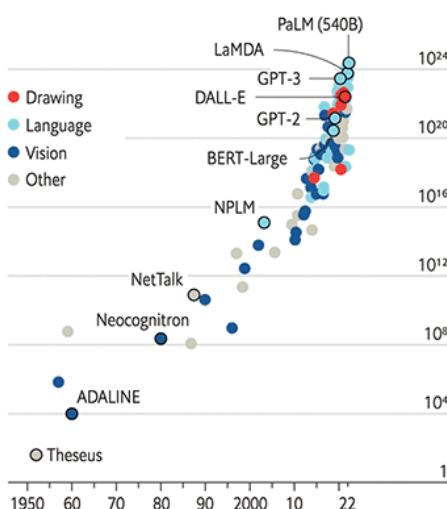
Generative AI has the same “why now” as AI more broadly: better models, more data, more compute. The category is changing faster than we can capture, but it’s worth recounting recent history in broad strokes to put the current moment in context.

Wave 1: Small models reign supreme (Pre-2015) 5+ years ago, small models are considered “state of the art” for understanding language. These small models excel at analytical tasks and become deployed for jobs from delivery time prediction to fraud classification. However, they are not expressive enough for general-purpose generative tasks. Generating human-level writing or code remains a pipe dream.

Wave 2: The race to scale (2015-Today) A landmark paper by Google Research ([Attention is All You Need](#)) describes a new neural network architecture for natural language understanding called transformers that can generate superior quality language models while being more parallelizable and requiring significantly less time to train. These models are few-shot learners and can be customized to specific domains relatively easily.

The blessings of scale

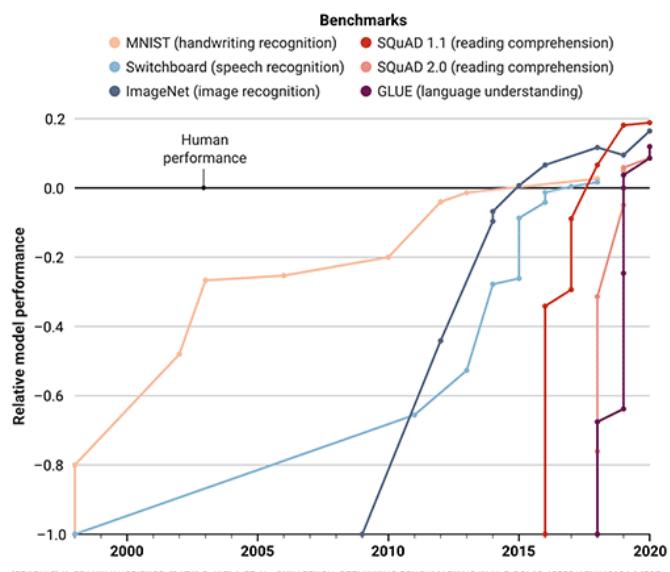
All training runs, estimated computing resources used
Floating-point operations, selected systems, by type, log scale



Sources: "Compute trends across three eras of machine learning", by J. Sevilla et al., arXiv, 2022; Our World in Data

Quick learners

The speed at which artificial intelligence models master benchmarks and surpass human baselines is accelerating. But they often fall short in the real world.



(GRAPHIC) K. FRANKLIN/SCIENCE; (DATA) D. KIELA ET AL., DYNABENCH: RETHINKING BENCHMARKING IN NLP, DOI:10.48550/ARXIV.2104.14337

As AI models have gotten progressively larger they have begun to surpass major human performance benchmarks.

Sources: © The Economist Newspaper Limited, London, June 11th 2022. All rights reserved;
[SCIENCE.ORG/CONTENT/ARTICLE/COMPUTERS-ACE-IQ-TESTS-STILL-MAKE-DUMB-MISTAKES-CAN-DIFFERENT-TESTS-HELP](#)

Sure enough, as the models get bigger and bigger, they begin to deliver human-level, and then superhuman results. Between 2015 and 2020, the compute used to train these models increases by 6 orders of magnitude and their results surpass human performance benchmarks in handwriting, speech and image recognition, reading comprehension and language understanding. OpenAI's GPT-3 stands out: the model's performance is a giant leap over GPT-2 and delivers tantalizing Twitter demos on tasks from code generation to snarky joke writing.

Despite all the fundamental research progress, these models are not widespread. They are large and difficult to run (requiring GPU orchestration), not broadly accessible (unavailable or closed beta only), and expensive to use as a cloud service. Despite these limitations, the earliest Generative AI applications begin to enter the fray.

Wave 3: Better, faster, cheaper (2022+) Compute gets cheaper. New techniques, like diffusion models, shrink down the costs required to train and run inference. The research community continues to develop better algorithms and larger models. Developer access expands from closed beta to open beta, or in some cases, open source.

For developers who had been starved of access to LLMs, the floodgates are now open for exploration and application development. Applications begin to bloom.



Illustration generated with Midjourney

Wave 4: Killer apps emerge (Now) With the platform layer solidifying, models continuing to get better/faster/cheaper, and model access trending to free and open source, the application layer is ripe for an explosion of creativity.

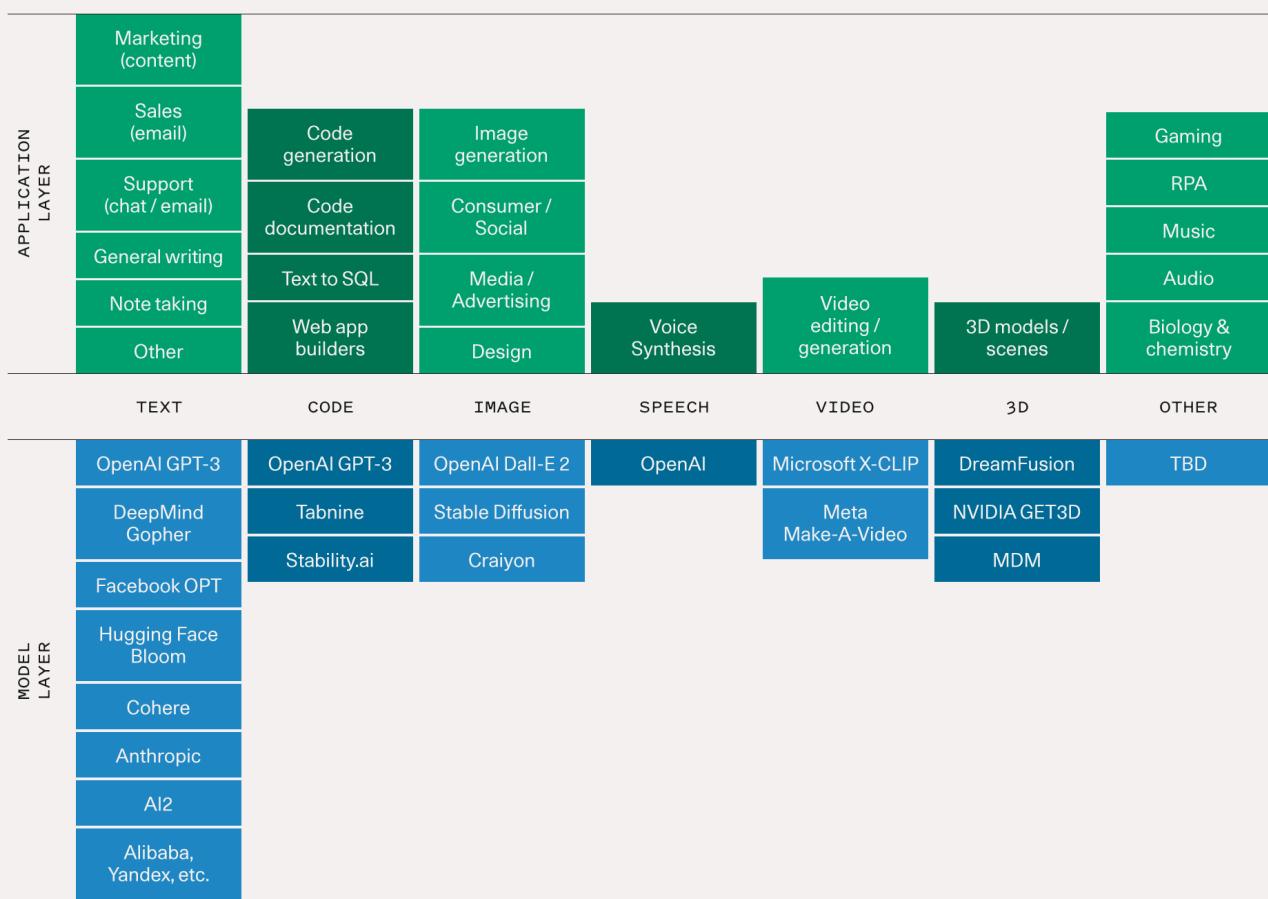
Just as mobile unleashed new types of applications through new capabilities like GPS, cameras and on-the-go connectivity, we expect these large models to motivate a new wave of generative AI applications. And just as the inflection point of mobile created a market opening for a handful of killer apps a decade ago, we expect killer apps to emerge for Generative AI. The race is on.

Market Landscape

Below is a schematic that describes the platform layer that will power each category and the potential types of applications that will be built on top.



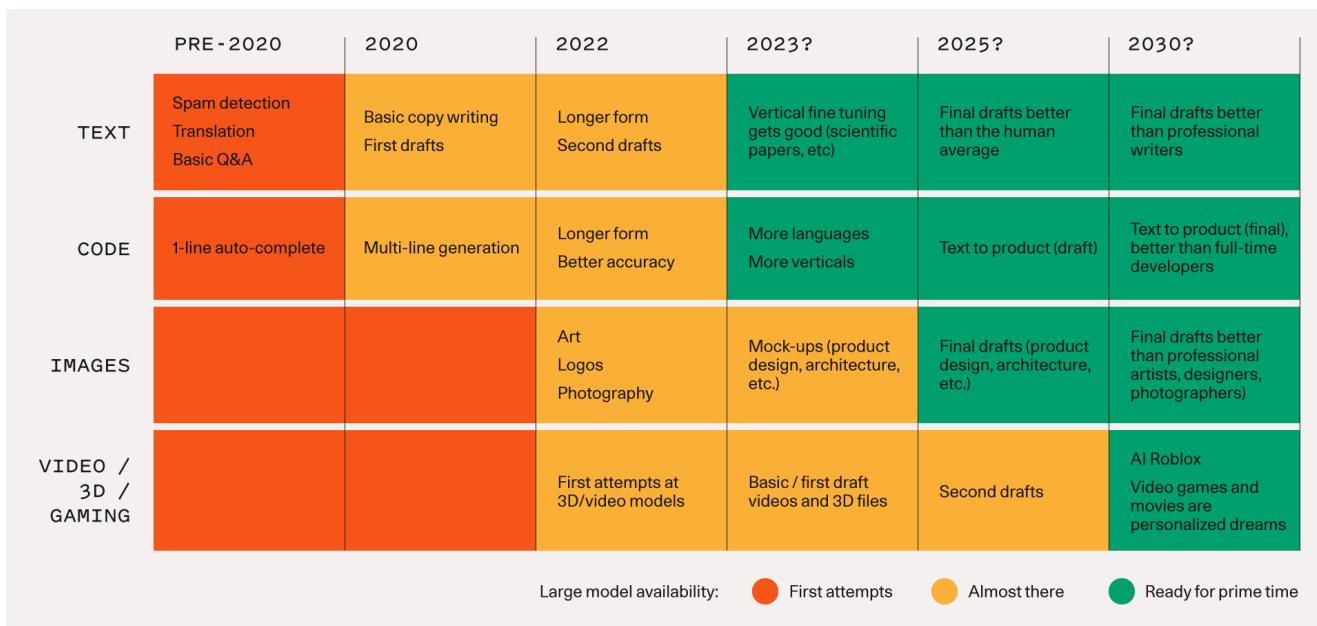
The Generative AI Application Landscape



Models

- **Text** is the most advanced domain. However, natural language is hard to get right, and quality matters. Today, the models are decently good at generic short/medium-form writing (but even so, they are typically used for iteration or first drafts). Over time, as the models get better, we should expect to see higher quality outputs, longer-form content, and better vertical-specific tuning.
- **Code generation** is likely to have a big impact on developer productivity in the near term as shown by GitHub CoPilot. It will also make the creative use of code more accessible to non developers.
- **Images** are a more recent phenomenon, but they have gone viral: it's much more fun to share generated images on Twitter than text! We are seeing the advent of image models with different aesthetic styles, and different techniques for editing and modifying generated images.
- **Speech synthesis** has been around for a while (hello Siri!) but consumer and enterprise applications are just getting good. For high-end applications like film and podcasts the bar is quite high for one-shot human quality speech that doesn't sound mechanical. But just like with images, today's models provide a starting point for further refinement or final output for utilitarian applications.
- **Video and 3D models** are coming up the curve quickly. People are excited about these models' potential to unlock large creative markets like cinema, gaming, VR, architecture and physical product design. The research orgs are releasing foundational 3D and video models as we speak.
- **Other domains:** There is fundamental model R&D happening in many fields, from audio and music to biology and chemistry (generative proteins and molecules, anyone?).

The below chart illustrates a timeline for how we might expect to see fundamental models progress and the associated applications that become possible. 2025 and beyond is just a guess.



Applications

Here are some of the applications we are excited about. There are far more than we have captured on this page, and we are enthralled by the creative applications that founders and developers are dreaming up.

- Copywriting:** The growing need for personalized web and email content to fuel sales and marketing strategies as well as customer support are perfect applications for language models. The short form and stylized nature of the verbiage combined with the time and cost pressures on these teams should drive demand for automated and augmented solutions.
- Vertical specific writing assistants:** Most writing assistants today are horizontal; we believe there is an opportunity to build much better generative applications for specific end markets, from legal contract writing to screenwriting. Product differentiation here is in the fine-tuning of the models and UX patterns for particular workflows.
- Code generation:** Current applications turbocharge developers and make them much more productive: GitHub Copilot is now generating nearly 40% of code in the projects where it is installed. But the even bigger opportunity may be opening up access to coding for consumers. Learning to prompt may become the ultimate high-level programming language.
- Art generation:** The entire world of art history and pop cultures is now encoded in these large models, allowing anyone to explore themes and styles at will that previously would have taken a lifetime to master.
- Gaming:** The dream is using natural language to create complex scenes or models that are riggable; that end state is probably a long way off, but there are more immediate options that are more actionable in the near term such as generating textures and skybox art.
- Media/Advertising:** Imagine the potential to automate agency work and optimize ad copy and creative on the fly for consumers. Great opportunities here for multi-modal generation that pairs sell messages with complementary visuals.
- Design:** Prototyping digital and physical products is a labor-intensive and iterative process. High-fidelity renderings from rough sketches and prompts are already a reality. As 3-D models become available the generative design process will extend up through manufacturing and production—text to object. Your next iPhone app or sneakers may be designed by a machine.
- Social media and digital communities:** Are there new ways of expressing ourselves using generative tools? New applications like Midjourney are creating new social experiences as consumers learn to create in public.

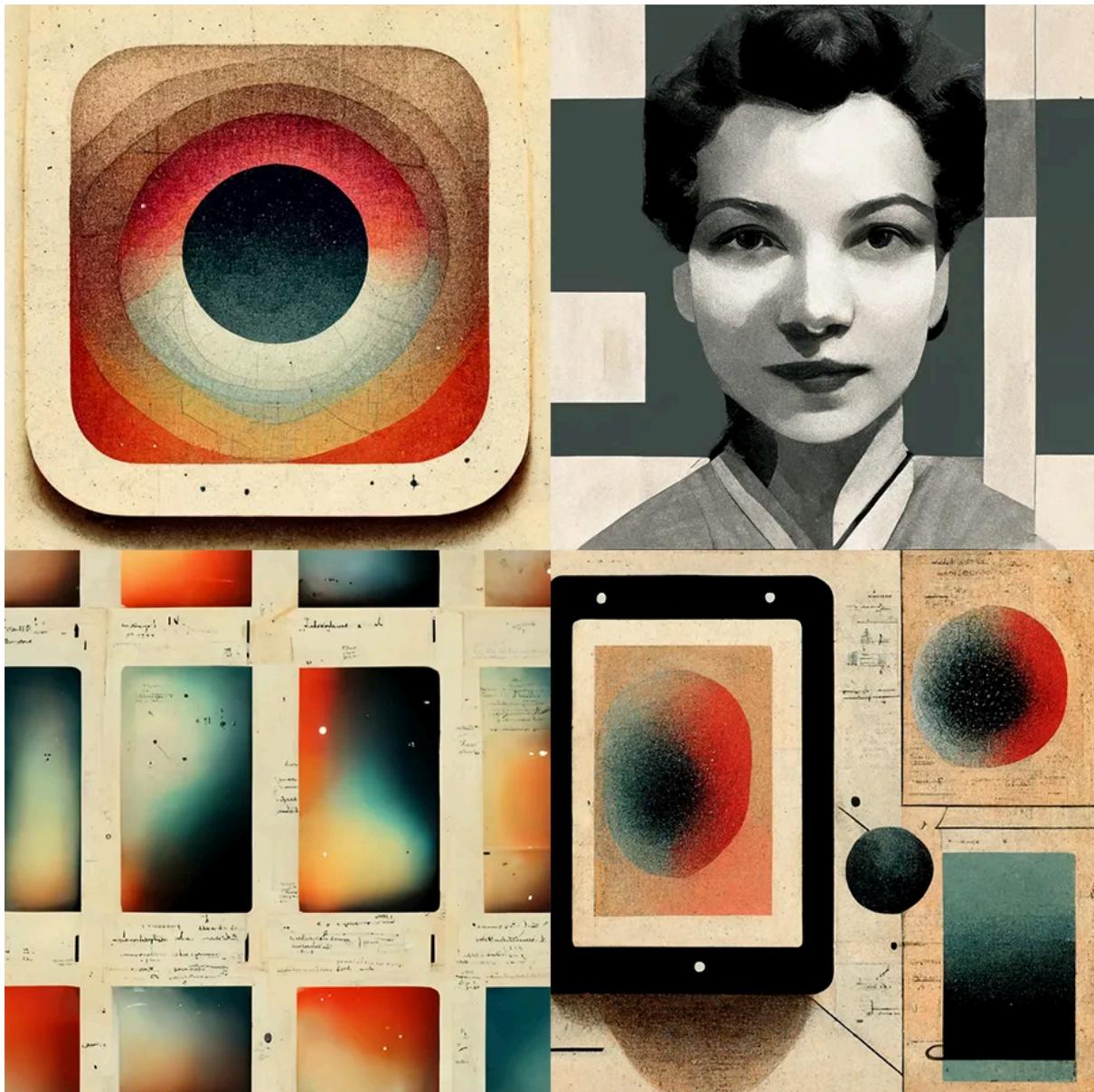


Illustration generated with Midjourney

Anatomy of a Generative AI Application

What will a generative AI application look like? Here are some predictions.

Intelligence and model fine-tuning

Generative AI apps are built on top of large models like GPT-3 or Stable Diffusion. As these applications get more user data, they can fine-tune their models to: 1) improve model quality/performance for their specific problem space and; 2) decrease model size/costs.

We can think of Generative AI apps as a UI layer and “little brain” that sits on top of the “big brain” that is the large general-purpose models.

Form Factor

Today, Generative AI apps largely exist as plugins in existing software ecosystems. Code completions happen in your IDE; image generations happen in Figma or Photoshop; even Discord bots are the vessel to inject generative AI into digital/social communities.

There are also a smaller number of standalone Generative AI web apps, such as Jasper and Copy.ai for copywriting, Runway for video editing, and Mem for note taking.

A plugin may be an effective wedge into bootstrapping your own application, and it may be a savvy way to surmount the chicken-and-egg problem of user data and model quality (you need distribution to get enough usage to improve your models; you need good models to attract users). We have seen this distribution strategy pay off in other market categories, like consumer/social.

Paradigm of Interaction

Today, most Generative AI demos are “one-and-done”: you offer an input, the machine spits out an output, and you can keep it or throw it away and try again. Increasingly, the models are becoming more iterative, where you can work with the outputs to modify, finesse, uplevel and generate variations.

Today, Generative AI outputs are being used as prototypes or first drafts. Applications are great at spitting out multiple different ideas to get the creative process going (e.g. different options for a logo or architectural design), and they are great at suggesting first drafts that need to be finessed by a user to reach the final state (e.g. blog posts or code autocompletions). As the models get smarter, partially off the back of user data, we should expect these drafts to get better and better and better, until they are good enough to use as the final product.

Sustained Category Leadership

The best Generative AI companies can generate a sustainable competitive advantage by executing relentlessly on the flywheel between user engagement/data and model performance. To win, teams have to get this flywheel going by 1) having exceptional user engagement → 2) turning more user engagement into better model performance (prompt improvements, model fine-tuning, user choices as labeled training data) → 3) using great model performance to drive more user growth and engagement. They will likely go into specific problem spaces (e.g., code, design, gaming) rather than trying to be everything to everyone. They will likely first integrate deeply into applications for leverage and distribution and later attempt to replace the incumbent applications with AI-native workflows. It will take time to build these applications the right way to accumulate users and data, but we believe the best ones will be durable and have a chance to become massive.

Hurdles and Risks

Despite Generative AI’s potential, there are plenty of kinks around business models and technology to iron out. Questions over important issues like copyright, trust & safety and costs are far from resolved.

Eyes Wide Open

Generative AI is still very early. The platform layer is just getting good, and the application space has barely gotten going.

To be clear, we don’t need large language models to write a Tolstoy novel to make good use of Generative AI. These models are good enough today to write first drafts of blog posts and generate prototypes of logos and product interfaces. There is a wealth of value creation that will happen in the near-to-medium-term.

This first wave of Generative AI applications resembles the mobile application landscape when the iPhone first came out—somewhat gimmicky and thin, with unclear competitive differentiation and business models. However, some of these applications provide an interesting glimpse into what the future may hold. Once you see a machine produce complex functioning code or brilliant images, it’s hard to imagine a future where machines don’t play a fundamental role in how we work and create.

If we allow ourselves to dream multiple decades out, then it’s easy to imagine a future where Generative AI is deeply embedded in how we work, create and play: memos that write themselves; 3D print anything you can imagine; go from text to Pixar film; Roblox-like gaming experiences that generate rich worlds as quickly as we can dream them up. While these experiences may seem like science fiction today, the rate of progress is incredibly high—we have gone from narrow language models to code auto-complete in several years—and if we continue along this rate of change and follow a “Large Model Moore’s Law,” then these far-fetched scenarios may just enter the realm of the possible.

Call for Startups

We are at the beginning of a platform shift in technology. We have already made a number of investments in this landscape and are galvanized by the ambitious founders building in this space.

If you are a founder and would like to meet, shoot us a note at sonya@sequoiacap.com and grady@sequoiacap.com.

We can’t wait to hear your story.

Ambitious founders can accelerate their path to success by applying to Arc, our catalyst for pre-seed and seed stage companies. Applications for Arc America Fall '23 are now open—apply [here](#).

PS: This piece was co-written with GPT-3. GPT-3 did not spit out the entire article, but it was responsible for combating writer's block, generating entire sentences and paragraphs of text, and brainstorming different use cases for generative AI. Writing this piece with GPT-3 was a nice taste of the human-computer co-creation interactions that may form the new normal. We also generated illustrations for this post with Midjourney, which was SO MUCH FUN!

Share

[Share this on Facebook](#) [Share this on Twitter](#) [Share this on LinkedIn](#) [Share this via email](#)

Related Topics

#AI

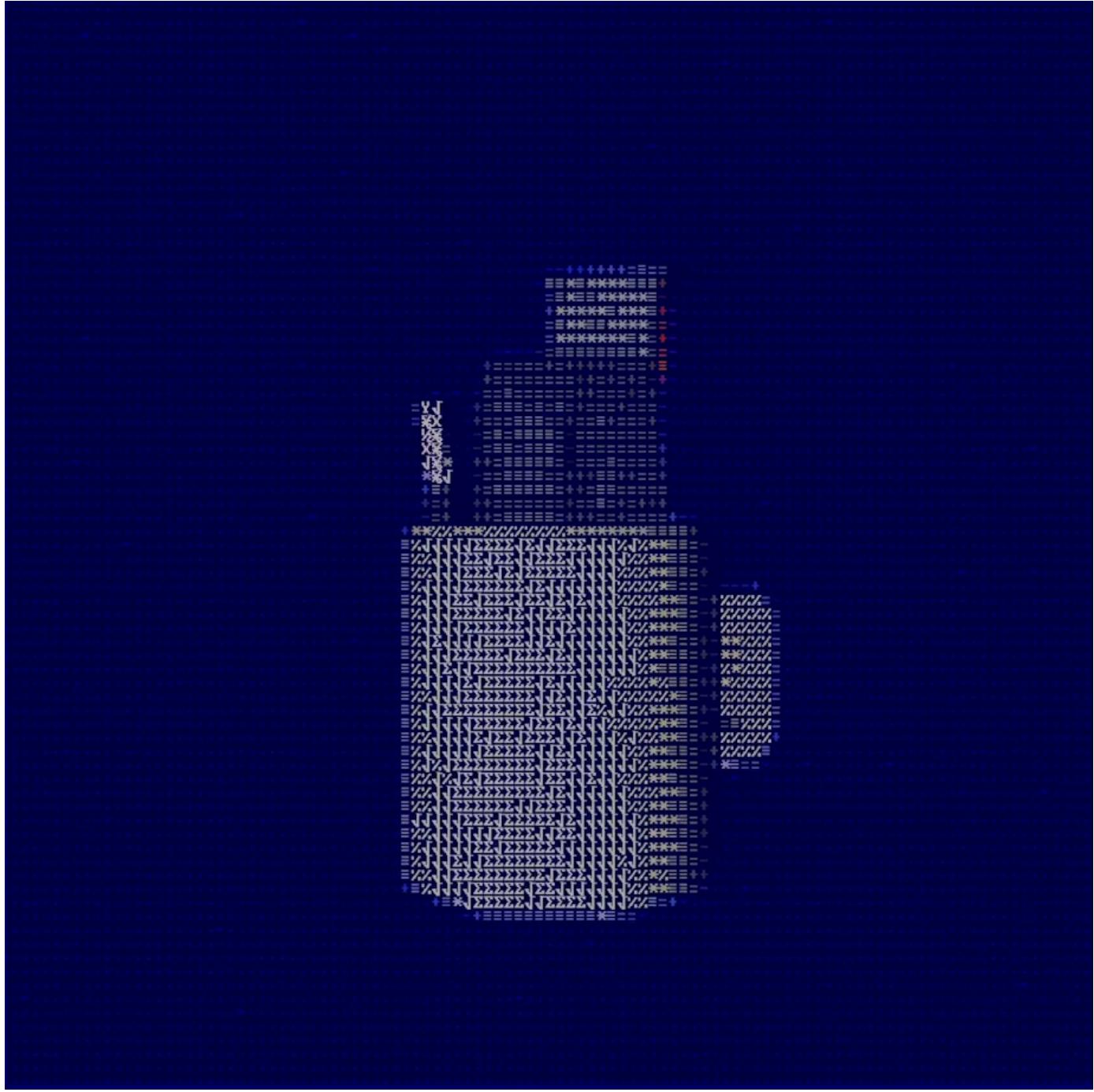


Developer Tools 2.0

[By Charlie Curnin, Josephine Chen and Stephanie Zhan](#)

[Perspective](#)

[Read](#)



AI Recruits a New Hybrid Workforce

[by Konstantine Buhler](#)

[Perspective](#)

[Read](#)

Partnering with Hugging Face: A Machine Learning Transformation

[By Pat Grady and Sonya Huang](#)

[News](#)

[Read](#)



[Partnering with Glean: Organizing Knowledge Work](#)

[News](#)
[Read](#)

[Join our Mailing List](#)

Get the best stories from the Sequoia community.

Email address

By navigating this website you agree to our cookie policy.