

Midjourney: A Product Strategy Case Study

Intro

Midjourney is an AI-driven platform that specialises in generating images from natural language inputs. It was built by an independent research lab with the same name, consisting of only 11 full-time staff to date, yet generating over \$200m ARR and 16 million users - having only started less than 2 years ago.

In this case study, we will examine the key factors contributing to its success, focusing on technical product and product marketing strategies, what lies ahead for the company, and some more high-level points for consideration.

I hope you enjoy the read.



Source: [Midjourney showcase](#)

The Beginnings of Midjourney: Foundation and Development

Early Days: The Formation by David Holz

Midjourney was established in 2022 by David Holz. Before that, Holz's background includes studies in physics and math, followed by a pursuit of a PhD in applied math, a period during which he also worked at NASA and the Max Planck Institute. In 2011, he relocated to San Francisco to start Leap Motion.

Leap Motion, known for its precise hand-tracking technology similar to a 3D mouse, gained significant traction with over 300,000 developers using the technology, attracted funding from

major VCs - an experience that would later influence Midjourney - and eventually was acquired by Ultrahaptics for \$30M in 2019.

Transition to Midjourney:

After Leap Motion, Holz found himself contemplating the future and what people would need in an uncertain world. He identified three core pillars he believed to be essential: reflection, imagination, and coordination. This contemplation and his personal philosophy significantly influenced the conceptualization of Midjourney.

Source: [Holz's origin story](#)

Origin of the name:

Holz attributes the concept of Midjourney to Daoist influence, specifically Zhuangzi. On the name itself though, Holz says he feels like we are actually *mid-journey* - that “we come from a rich and beautiful past, but ahead is this wild, unimaginable, unfathomable future”.

Source: [Reddit](#)

The Vision is bigger than you think:

Holz envisioned Midjourney not just as a technology product but as a platform for creativity and expression. He focused on creating an AI-powered platform that would transform text prompts into visual imagery, enabling users to explore new realms of creativity and imagination. This vision was rooted in his belief in the power of AI to expand human imagination and capability.

- “*We don’t think it’s really about art or making deepfakes, but — how do we expand the imaginative powers of the human species?*” Source: [Verge Interview](#)

Right now, there is a lot of fear-mongering around AI, but Holz sees things differently, he likens AI to an engine - engines are without feeling, without motivation, and without direction - it is only the human’s application of the engine that those things are derived.

An engine is a tool that is to be wielded by humans, to get us from A to B much more efficiently, and in doing so, take us on a journey that opens our minds to new ways of thinking and looking at the world. Midjourney was designed to help influence new frontiers of imagination, to be a creative partner, not a replacement.

In an interview with The Verge, Holz goes on to compare the discovery of these AI models to the discovery of water. In the sense that they both represent fundamental elements that can be harnessed for transformative purposes. Just as water can be both a peril and a boon to

humanity – capable of causing harm but also essential for life and progress – AI too holds dual potentials. The point is not to be afraid of its potential, but to understand how we can build tools to harness its potential to better our lives:

- “How do we teach people to swim? How do we make boats? How do we dam it up? How do we go from people who are scared of drowning to kids in the future who are surfing the wave? We’re making surfboards rather than making water. And I think there’s something profound about that.”

Opinion:

- I don't think the importance of the vision of a company gets as much emphasis as it should. Perhaps we forget that humans are driven by emotion, and when you develop a moonshot, something that is a little bit abstract, but is exciting and adds value to the world - *supercharging human imagination* in this case - that is what galvanizes the team around you and what attracts the best talent to you.

The Ethos

The ethos of Midjourney, as encapsulated in the quote, “It’s just about having a home for the next 10 years to work on cool projects that matter —hopefully not just to me but to the world — and to have fun,” reveals a company culture deeply committed to passion-driven innovation and the pursuit of projects with global impact. Source: [Verge Interview](#)

This eleven-person team operates without external capital, a unique position that liberates them from the typical financial pressures and constraints often faced by startups. This independence is key to understanding their approach. The absence of external financial motivations and external forces trying to pull the strings, allows for a purer focus on crafting a product that genuinely connects with users, fostering a deeper level of engagement and satisfaction.

Bottomline is Holz has managed to protect the core of the business and his team from being pulled in the wrong direction, optimising for a product that users love - nothing more, nothing less.

Opinion:

- I think this is an important lesson for early stage founders. Whenever we read about startups in the media, it's usually about how much money the startup raised **and therefore** how much it is valued at. So in our minds we associate the amount of money you can raise to the amount of value a startup creates.
 - i.e. we assume:

- value a startup creates == valuation of the company == $f(\text{amount of money raised from VCs})$
- I think founders, and perhaps even VCs, really need to disconnect from this mental model, and realise that the true value of a company, is a function of the value created for each individual user of your product. And when you optimise for this, as Midjourney clearly do, the financial rewards will come - they were always a lagging indicator anyway.



Product Strategy:

It's worth looking at their product strategy from two angles: the technical product strategy, i.e. more focused on how they were building and optimising their model, and the product marketing strategy, more focused on how they positioned and their product and engaged with users.

An intro to product:

Midjourney, like other generative AI platforms, operates on a foundation of advanced machine learning techniques, primarily using diffusion models. The core principle behind these models is to start with a sample image, incrementally add noise, and then train the model to reverse this process, effectively learning to generate new images that are similar to the original. This process allows for the creation of unique, yet familiar, visual content.

The training of these models is a data-intensive process. Platforms like Midjourney typically gather vast amounts of data from the internet, which includes scraping images and text. For instance, Midjourney has utilized open-source training models and extensive datasets, such as the 2 billion image-text pairs from the English subset of CLIP's open dataset, created by the German non-profit LAION. This approach of aggregating and processing massive datasets enables these AI models to learn and replicate a wide range of styles and content, making them versatile tools for various creative applications. By continuously analyzing user interactions and

preferences, platforms like Midjourney further refine their models, ensuring that the generated content resonates with user expectations and emerging trends.

Technical Product Strategy

Offense - Default style of MJ:

Midjourney's default style is intentionally designed to be more artistic and interpretative than the specific input provided by the user, serving a distinct purpose in its product strategy. According to Holz, the rationale behind this approach is that users often don't know exactly what they want when generating images. If precise replication were the goal, one could simply use Google Images. However, platforms like Midjourney aim to elevate human imagination, necessitating a more creative and proactive role in the image generation process.

The essence of Midjourney's approach is to avoid the mundane. For instance, a simple prompt like "dog" could yield a straightforward photo, but that lacks creativity and purpose in the context of AI-driven artistry. Instead, Midjourney aspires to produce works that are not just responses to prompts but artistic interpretations. This is evident in the model's inclination towards whimsical, abstract, and somewhat peculiar outputs, often blending elements in unexpected yet aesthetically pleasing ways. Holz notes that the model has preferences, such as favoring certain colors and styles, which contributes to its unique artistic identity.

The style is often described as "a bit whimsical and abstract and weird," noting that it blends elements in unexpected and aesthetically pleasing ways, often favoring certain colors and themes. This distinctive style is a deliberate choice, ensuring that the output is more than just an answer to a query - it's a creative journey, offering users a blend of beauty, surprise, and artistic flair.





Source: [Miss Journey](#) - a default face the model has a tendency to draw

Defense: On restrictions

What's important about Midjourney's product strategy, is not just what it allows users to do, but also what it does not allow users to do.

Another important part of Midjourney's product strategy is not only the creative possibilities of what users can do, but also the limitations it places on what users cannot do. Considering the boundless scope of user imagination, generative AI platforms have sometimes been used to create content that is graphic and violent.

One way Holz and his team found a way to deal with this, was to inject accountability of the images created by putting that user's name on the images created - "When you put someone's name on all the pictures they make, they're much more regimented in how they use it. That helps a lot." Essentially, by promoting transparency over who made it, it almost created a self-policing mechanism.

On top of this, they added in more robust guard rails, from moderators to the team actively intervening and banning specific words such as "ultragore and everything within a mile of that"

Opinion:

- *I think the point here is that yes there's this moonshot vision that's clear to the team and the users, but Midjourney also set up guardrails to ensure that no one veers off from the track, and ends up morphing the platform into something the team did not want it to be.*

On artists:

The issue of copyright within platforms like Midjourney and ChatGPT are still in open discussion right now, so I won't delve into this too much.

But it is worth noting that the fact that Holz and his team have been aware of addressing the concerns of the artistic community from the start, has likely also played a role in their success. This is not just to do with anticipating liability issues, but for a product that is built on a community, the community needs to maintain trust in the product, and by actively engaging with their artists it is no wonder why around 4 million of its users are working professional artists.

- "We do have a lot of artists in the community, and I'd say they're universally positive about the tool, and they think it's gonna make them much more productive and improve their lives a lot. And we are constantly talking to them and asking, "Are you okay? Do you feel good about this?" We also do these office hours where I'll sit on voice for four hours with like 1,000 people and just answer questions."

Source: [Verge Interview](#)

Product Marketing Strategy

Partnership with Discord

Their partnership with Discord has more to it than at first glance.

Firstly, by not being on a simple to access website, Discord in a way, acted as the sandbags, holding back the flood of users that inevitably come with virality - a problem that OpenAI faced. In fact, it allowed the team to continue to engage with the community they did have, and perfect their product, before opening the flood gates - which they are now doing.

Secondly, even in its very early stages, Midjourney still had to support hundreds of thousands, and soon millions, of users trying to access their model. By piggybacking off of Discord's infrastructure, to handle the traffic, Midjourney was able to keep its head well above the water, and is also a contributing factor to why the team could stay as small as it has.

Third is engagement. It's not simply just that Discord already had a large and active community, but also how the engagement specifically ended up influencing their output. The idea is essentially a "round-robin" story - where one person starts, and then another person adds to it, and then another, and another, and by the end, you create something that no one individual could have dreamed of.

Holz explained it pretty well in his interview with the Verge:

- "We found very quickly that most people don't know what they want. You say: "Here's a machine you can imagine anything with it — what do you want?" And they go: "dog." And you go "really?" and they go "pink dog." So you give them a picture of a dog, and they go "okay" and then go do something else.
- Whereas if you put them in a group, they'll go "dog" and someone else will go "space dog" and someone else will go "Aztec space dog," and then all of a sudden, people understand the possibilities, and you're creating this augmented imagination — an environment where people can learn and play with this new capacity."

Furthermore, because it is so community-driven, you automatically end up with art that is fun, diverse and completely original. And because you're showcasing the generated images back into the community, there is a much higher chance of things going viral - and when they do, it reinforces interest in the Midjourney platform - case and point the Pope in a puffer jacket image that **broke the internet**, and led to even more mass interest in Midjourney.

Opinion:

- I think this point here opens up a whole new can of worms. Because yes it AI can augment human output, but humans adjusting the output from another human's AI output, creates a flywheel that is way beyond what any one person, or even isolated interaction with AI can achieve. And honestly, its this that is the future of AI - human + AI + more human inputs creating a flywheel of innovation.

Sandbox-and-watch strategy

Perhaps one of their main product strategy could be described as sandbox-and-watch. By this I mean, they created the playground of what could be done, put in some general guidelines of what should not be done, and then just watch the community take the product in its different directions.

Whilst this has led to Midjourney being used to create some incredibly fun art (an anticipated outcome), as well as using it for bad (also anticipated), some users have been using it for art

therapy - where people create images of their loved ones who recently passed away - definitely not anticipated by the team.

And it's not just a few users, around 20% of all users on Midjourney use it for art therapy. In fact the man behind the viral **Pope in a puffer jacket** initially started using the platform to create images of his brother who passed away.



Source: [Buzzfeed](#)

Opinion:

- I think there is something to this strategy - because by simply providing the sandbox, you are stepping away from any confirmation bias you/team has (because when you build something, you just assume it will be used in the way you're anticipating), and instead, you open the door to diverse and unforeseen user innovations. This not only challenges your initial assumptions but also enriches the product's evolution, driven by actual user creativity and need.

Business model

A few brief points worth noting about the business: From a top-line perspective, Midjourney are currently doing around \$200m ARR, and for a team of only 11, that's pretty impressive. It has around 16million users, with 30% of their users being professionals, which likely includes

industries like graphic design, marketing, and perhaps even entertainment. How exactly the outcomes of the legal proceedings around generative AI will impact Midjourney, and this segment of its customers, is still yet to be seen.

On the cost side, the expense of training image models is significant, around \$50,000 per training session, especially considering that multiple iterations are often necessary to achieve the desired accuracy and quality of the models. This iterative process, which might require “three tries or 10 tries or 20 tries”, implies a considerable investment in research and development. “It is expensive. It’s more than what most universities could spend, but it’s not so expensive that you need a billion dollars or a supercomputer.” They are also running on \$20,000 servers, which they rent. The point here is that the compute power is enormous for simply just generating the images, in the order of thousands of trillions of operations (petaops) per second - “there has never been a service before where a regular person is using this much compute”. Nonetheless, Holz anticipates costs will drop as competition increases and investors plough in more money.

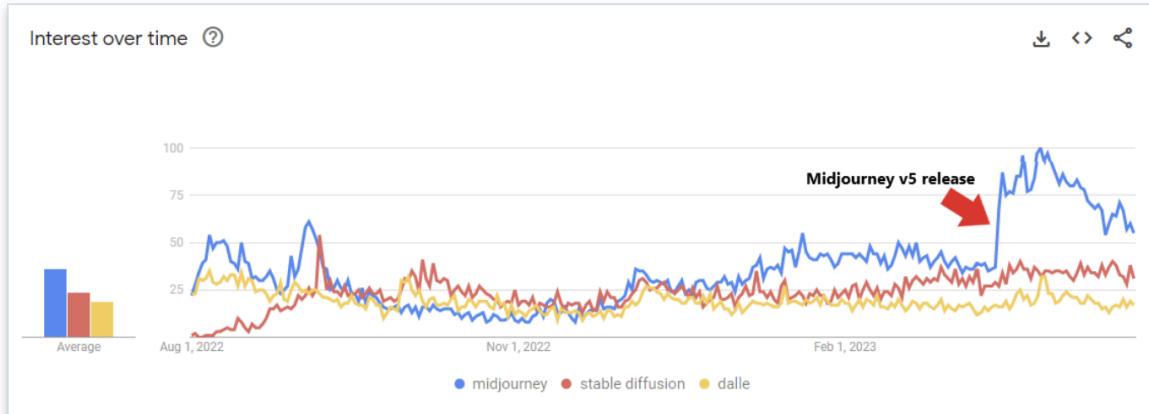
Opportunities

This month, Jan 2024, Midjourney have released their platform onto their own web service in an effort to increase accessibility and stay competitive. Though it's worth noting, they are still sticking to the strategy of holding back the flood until they are ready as the website will initially be available only to people who've racked up more than 10,000 images via Midjourney on Discord - allowing them to test and refine the platform's experience.

The next step beyond image generation is obviously video generation. Many platforms are already making notable strides in video generation technology. As this field evolves, the platform that excels in producing high-quality video content is poised to gain a substantial competitive edge.

The ability to generate videos effectively and innovatively could become a crucial determinant in leading the market. This suggests that the future of AI in visual media might very well hinge on mastering video generation. It's likely that the increased revenues from greater accessibility via the web platform, will help fund the training of the models for video enhancement.





Source: Google Trends (Popularity Index, 100 = Peak)

Threats: Competition:

There are a number of similar platforms to Midjourney, from OpenAI's DALLE to Stable Diffusion. The graph above shows that pre-V5 release, the gap between these three platforms has been very tight, and it's likely that this will continue to be an arms race between the top players.

What's important here, is that whilst objectively they all seem to do the same thing, generate new, creative images, they still do it in very different ways. The graph below is from a quantitative study analysing the performance of the 3 different models against real images. FID (Fréchet Inception Distance) is simply a way to evaluate the quality of images generated by models - where the lower the FID score, the more realistic the image.

As you can see, Stable Diffusion is much better at generating hyper-realistic images, but the importance of this completely depends on who's using it and for what. Midjourney has a core artistic community, likely because the Midjourney is not so hyper-realistic, with its default style, it provides a better use case for continuing to "expand the imaginative powers of the human species".



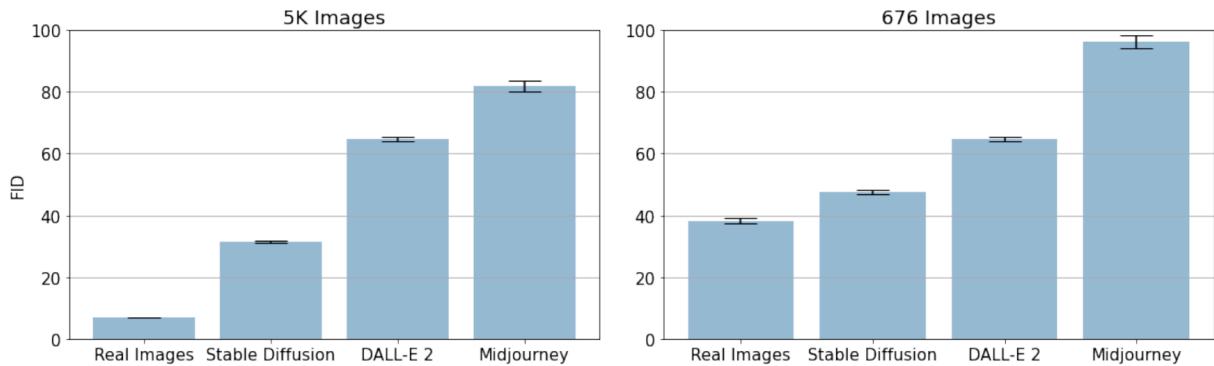


Figure 1: Left: FID scores of models over random sets of 5000 faces. Right: Results with samples of size 676 per model. Notice that the lower the FID, the better. Results are averaged over 10 runs.

Source: [Quantitative Comparison Study](#)

OTSOG Media
Case #1



OTSOG[©]