A Netflix Special:

Product Case Study (Tech Focused)

Contents:

- Snapshot
- Background
- Defence: The Simian Army
- Midfield: AWS and Infrastructure
- Offense: Personalisation / Search and R+D

Snapshot

Netflix has transitioned from a media-logistics business to a media-tech giant, with the sole purpose of giving their users the ability to watch what they want, when they want. In this case study, we analyse how Netflix were able to leverage technologies to build such a dominant product on three levels: defence (how it meets the bare minimum of user expectations), midfield (how its infrastructure allows it to assist in meeting user expectations and facilitate innovation), offense (how Netflix's innovations and leveraging of data allow it to further its competitive moat):

The Defence: The Simian Army represents Netflix's defensive strategies, focusing on resilience and system robustness. By deliberately introducing failures into their system, Netflix enhances its ability to withstand real-world issues, ensuring high availability and reliability for users.

The Midfield: AWS and cloud infrastructure are akin to the midfield, serving as the backbone that supports both defensive and offensive plays. This infrastructure allows Netflix to scale efficiently, handle massive amounts of data, and maintain operational flexibility, enabling rapid deployment and innovation.

The Offense: Personalization, search, and R&D are akin to some of Netflix's offensive strategies, as they are crucial for strengthening its competitive edge. Through advanced algorithms and continuous experimentation, Netflix tailors content to individual preferences, enhancing user engagement and satisfaction. This forward-looking approach drives growth and sets Netflix apart from competitors.

Background

Netflix is the leading streaming entertainment service, with around 44% market share, offering a wide array of TV shows, movies, anime, documentaries, and more across a variety of genres and languages. Founded in 1997 by Reed Hastings and Marc Randolph in Scotts Valley, California, it initially started as a DVD rental service by mail. However, Netflix pivoted to streaming in 2007, marking the beginning of its transformation into a global streaming powerhouse.

Today, Netflix operates in over 190 countries, providing millions of subscribers with content that can be watched virtually anywhere, on any internet-connected screen. Known for its significant investment in original content, Netflix has produced critically acclaimed series and films, further solidifying its position as a leader in the entertainment industry.

Netflix employs around 12,000 staff, with two-thirds being technical. Despite transitioning from a media-logistics service, to a media-tech company, at its heart it still serves the same purpose: give their users the ability to watch what they want, when they want. Technology just enhanced that service offering, whilst finding new ways to create more value for each user.

> "At Netflix, we want to entertain the world. Whatever your taste, and no matter where you live, we give you access to best-in-class TV series, documentaries, feature films and games. Our members control what they want to watch, when they want it, in one simple subscription."
> - Netflix

Throughout the OTSOG case series so far, we tend to analyse companies from both a technical-product and product-marketing lens. For this Netflix Special however, we've decided to double-down on the technical angle, simply because there are a number of learning points on this angle alone that are worth an in-depth exploration.

To simply matters further, we've split the technical strategies into 3 parts, as mentioned in the snapshot: the defence, midfield and offense. Hopefully, this does not come across as an oversimilpfication, but we believe that it helps provide a clearer framework to understand the multifaceted approach Netflix employs to innovate, secure, and expand its platform effectively.

The strategies discussed here are by no means an exhaustive list, but they illustrate key elements of how Netflix leverages technology to maintain its industry leadership and innovate within the streaming landscape.

I hope you enjoy the read.

The Defence: The Simian Army

Arguably, a good metric of the capabilities of a company's tech team, is their ability to build and ship open-source, developer-centric products, that are widely used, but not known to the average consumer. Not only do such contributions showcase technical ability and thought leadership, but they also highlight a commitment to advancing the technology ecosystem: Facebook built React (one of the most widely-used JavaScript libraries), LinkedIn built Kafka (a real-time data-streaming platform), Google built Kubernetes (a system for automating and deploying containerised applications), and Netflix built Chaos Monkey and subsequently the Simian Army.

Chaos Monkey is a tool that intentionally introduces failures into the infrastructure to test its resilience and reliability. It works by disabling computers in Netflix's cloud production network to test how remaining systems respond to the outage, thereby ensuring that Netflix's services can tolerate the failure without any customer impact - "we have to be stronger than our weakest link" - Director of Cloud & Systems Infrastructure at Netflix, Yury Izrailevsky.

The Netflix Cloud team describe its philosophy as this:

"Imagine getting a flat tire. Even if you have a spare tire in your trunk, do you know if it is inflated? Do you have the tools to change it? And, most importantly, do you remember how to do it right? One way to make sure you can deal with a flat tire on the freeway, in the rain, in the middle of the night is to poke a hole in your tire once a week in your driveway on a Sunday afternoon and go through the drill of replacing it. This is expensive and time-consuming in the real world, but can be (almost) free and automated in the cloud."

"By running Chaos Monkey in the middle of a business day, in a carefully monitored environment with engineers standing by to address any problems, we can still learn the lessons about the weaknesses of our system, and build automatic recovery mechanisms to deal with them. So next time an instance fails at 3 am on a Sunday, we won't even notice."

From the success of the Chaos Monkey, the Simian Army was born. This was a suite of different types of chaos monkeys, that would introduce different kinds of chaos and hence test different vulnerabilities, assessing Netflix's ability to survive them - similar to how the human body has different types of immune system cells, but all still serving the same purpose, to ensure the health of the host. The army includes:

- Latency Monkey - introduces artificial delays, making them intentionally slow, to see if the serves still work well when they are requests to the server are delayed.
- Conformity Monkey - finds instances that don't adhere to Netflix's best practices and shuts them down.

- Doctor Monkey - checks if our services are healthy by looking at their regular health checks and other signs, like how busy the CPU is. If it finds services that aren't doing well, it takes them out of use.
- Janitor Monkey - searches for unused resources and disposes of them to ensure the cloud environment is running free of clutter.
- Security Monkey - finds security violations or vulnerabilities, and terminates those instances.
- Chaos Gorilla and many more

Netflix's strategy of deliberately introducing problems into their systems to uncover weaknesses early has revolutionized the way companies ensure the robustness of their distributed systems. This approach, known as chaos engineering, has established a new standard for testing system resilience, inspiring a shift across the tech industry towards more reliable and fault-tolerant software development practices.

By identifying and addressing vulnerabilities before they escalate, the Simian Army fortifies Netflix's defence on service delivery, ensuring Netflix can continue to do the thing that customers are paying it for - let them watch the content they want, when they want it, without failure.

Figure 1 - A pictographic of Netflix's Simian Army, source.

The Midfield: AWS and Infrastructure

"We serve hundreds of thousands of requests per second..we serve over 1 billion of hours of content per month..and all of this is being run out of AWS infrastructure"
- Eva Tse, Director of Big Data Platform, Netflix

Amazon Web Services (AWS) is a comprehensive cloud computing platform, providing a wide range of services like computing power, database storage, and content delivery. It's known for its scalability, reliability, and flexibility, making it a popular choice for businesses of all sizes. What makes the use of AWS for Netflix even more interesting, is that Amazon Video is a direct competitor to Netflix, so why use it?

Uprooting your set infrastructure is often driven by necessity. For Netflix, this necessity came from two catalysts:

- First, an initial wake-up call - a two-day outage of their streaming services
- Second, the uptick in streaming demand - which provided the momentum to see them through the transition.

Before 2008, the strategies and practices for handling system failures in software development were mainly focused on traditional IT infrastructure, i.e., dedicated hardware and in-house data centres. Developers and engineers employed methods like fault tolerance and redundancy, but the widespread adoption of cloud computing and distributed systems was still very nascent.

Netflix started to rethink this following a 2-day outage to their streaming service, due to a hardware failure (specifically their data storage infrastructure) in 2008. The outage underscored the limitations of traditional, physical infrastructure in meeting the demands for continuous service availability. Netflix realized that to achieve the level of resilience and operational agility required, they needed an infrastructure that could dynamically scale and recover from failures seamlessly.

Cloud computing, with its promise of high availability, scalability, and cost-efficiency, offered a solution. AWS, known for its robust cloud services, presented an opportunity for Netflix to leverage these cloud benefits. By adopting AWS, Netflix could focus on improving and expanding its service offerings without the constraints of managing physical hardware, thus ensuring that its streaming service could reliably meet the rapidly growing demand from users worldwide.

But another existential threat was also needed to push them through the transition. In 2009, for the first time, streaming demand overtook their DVD rental demand, meaning they needed a vast increase in their data centre capacity. This is important because, like many companies at the time, their IT infrastructure was scaled to how many employees they have (i.e. internal

company demand-driven), but, especially for customer-facing companies like Netflix, it needed to scale to how many customers they have.

This shift from DVD to streaming is the crux of the need for AWS. Because for any one interaction between a user and the movie they watch, Netflix had a 100-fold increase in the amount of traffic to their data centre per movie/TV show (things like storing where you stopped, what shows you've seen etc); and combined with a 10-fold increase in views per week, Netflix now had a 1000-fold increase in traffic to their datacenter.

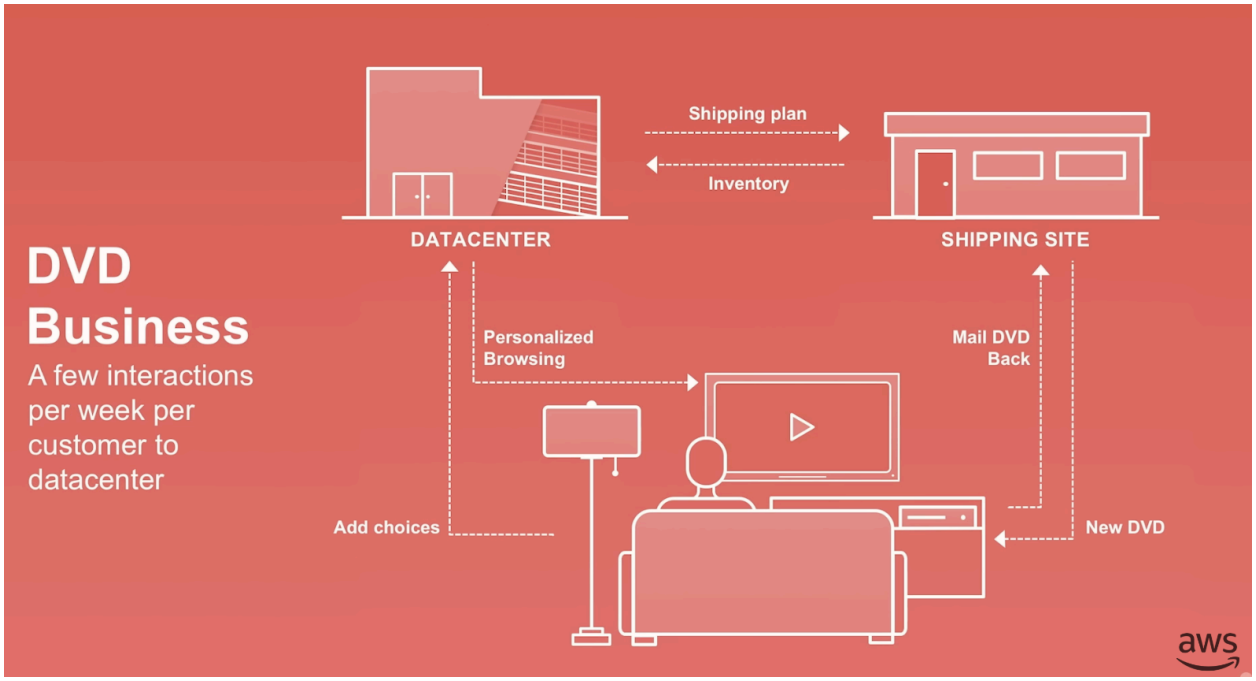


Figure 2 - The image presents the logistics of a DVD rental business, showing limited customer interactions with a data centre for personalized browsing and a physical process for mailing DVDs to and from a shipping site.

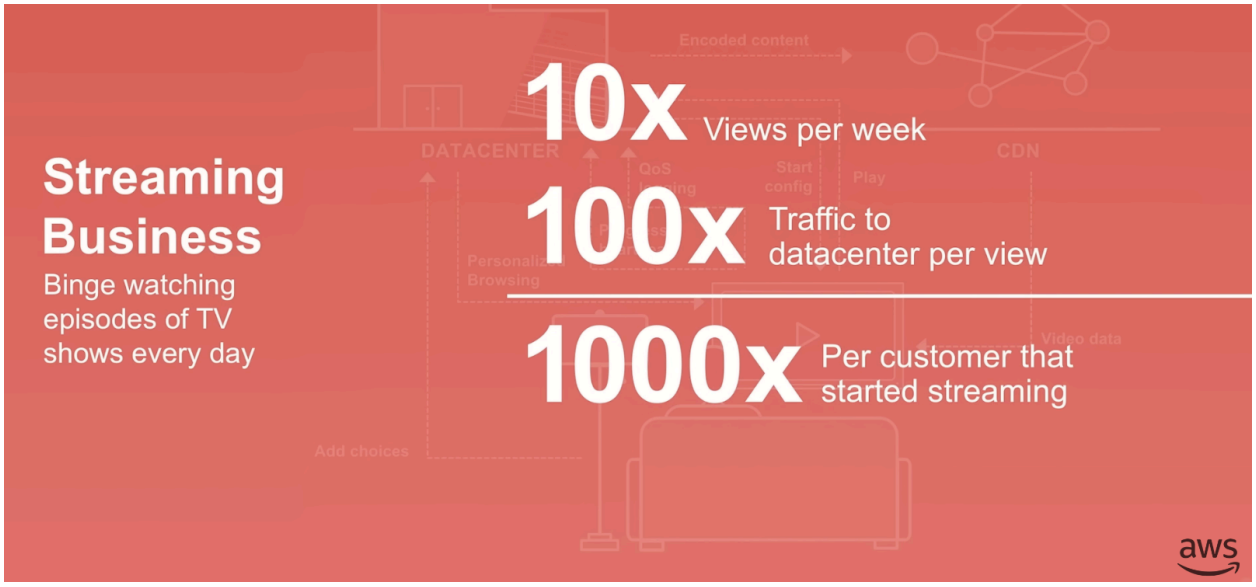


Figure 3 - The image outlines the exponential growth in a streaming business, highlighting a 10x increase in weekly views, 100x more data centre traffic per view, and a 1000x rise in traffic with each customer shift to streaming.

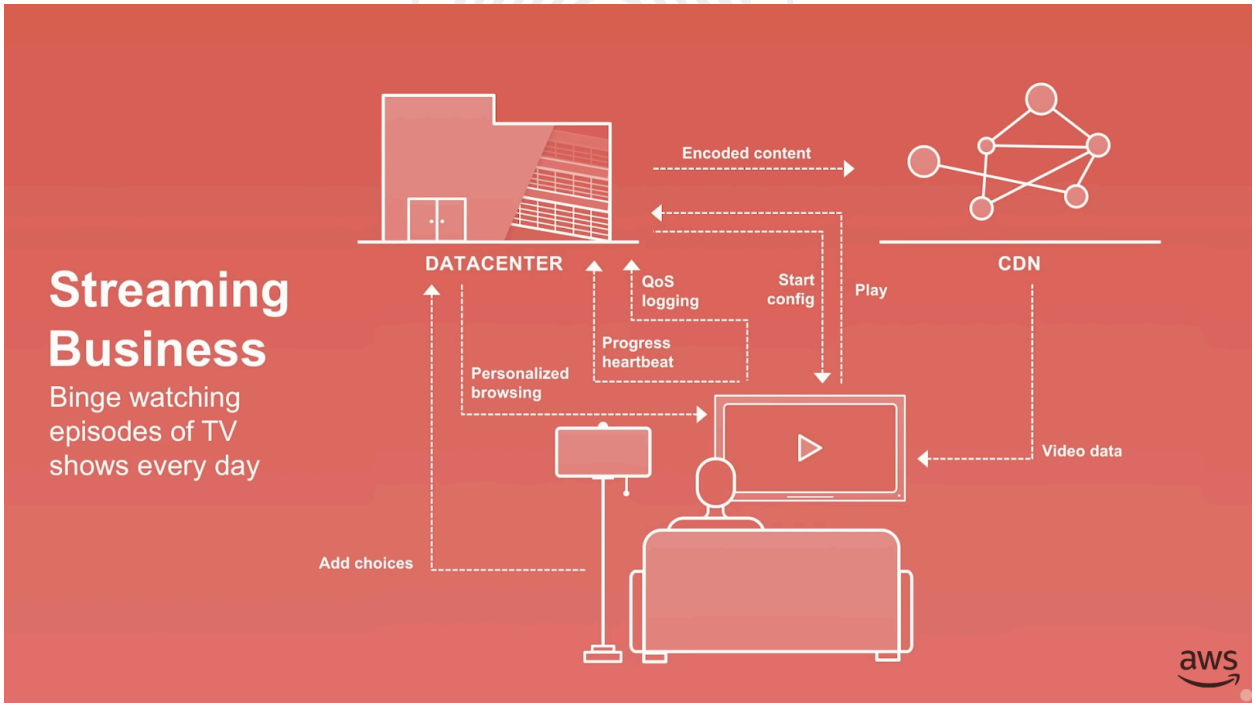


Figure 4 - The diagram shows the workflow of a streaming service with a data centre processing personalized content and a CDN distributing encoded video to users. It highlights infrastructure components for delivering a seamless binge-watching experience.

To figure out how to build data centres fast and keep up with the demand, Netflix had two options:

- (1) recruit a data centre operations team, try to guess how much capacity they would need and build it before it was needed (massive upfront costs).
- (2) use the elastic compute services of AWS (EC2) which scaled with the demand needed, despite getting into bed with one of their biggest competitors (Prime Video).

After a meeting with Amazon and an understanding that AWS and Prime Video were completely separate businesses, Netflix was satisfied enough to build with AWS and hasn't looked back since.

A few case examples of the core benefits of AWS for Netflix:

Cost efficiency:

With AWS's pay-as-you-go model, Netflix pays only for the compute time that is actually used. This model allows for billing by the hour or even by the second, depending on the instance type, providing flexibility in resource utilization. As a result, tasks can be processed faster by using more machines for a shorter time without increasing the total compute hours and, consequently, the cost.

To illustrate: Encoding movies and shows (i.e. optimising raw video files for online streaming) requires substantial compute power, especially as the volume of content grows. AWS's pricing structure (where 10 instances for 10 hours costs the same as 100 instances for 1 hour) enables Netflix to scale up their compute capacity as needed to swiftly manage large backlogs of shows. This capability allows for rapid content availability to customers without incurring additional costs.

Speed of iterations:

Before the adoption of cloud computing, Netflix software engineers had to request additional physical servers from the IT department to conduct tests and experiments, a process that could take days or even weeks. This separation of servers was necessary to maintain isolation and prevent any experimental changes from impacting the live product.

However, with AWS, engineers can now launch virtual instances for their experiments almost instantaneously—within seconds or minutes. This capability greatly enhances their flexibility, allowing them to run more experiments at a faster pace, thereby significantly speeding up Netflix's innovation cycle (something we cover in more detail under *The Offense*).

New ways of working in the cloud:

During the production of "The Crown" Season 4, the series encountered the daunting task of generating over 600 visual effects (VFX) shots amid the global pandemic constraints.

The transition to a cloud-based workflow on AWS empowered a relatively small in-house VFX team of 10 artists to work remotely and efficiently, leading to the delivery of the 10-episode series in a mere eight months.

"Our users expect that when they turn on the Netflix service, that it will just work. And that is partly due to the fact that we're on AWS Cloud"
- Eva Tse, Director of Big Data Platform, Netflix

As you can see, AWS allowed Netflix to orchestrate their defence and attack strategies:
- Defensively, AWS enabled them to not only maintain their core service offering — allowing users to watch content on-demand — but also to enhance its robustness and scalability.
- Offensively, AWS facilitated much faster experimentation and iteration rates, allowing Netflix to develop new features that increased stickiness and attracted more subscribers to the platform.

The Offense: Personalisation, Search and R+D

Personalisation

"Personalization is one of the pillars of Netflix because it allows each member to have a different view of our content that adapts to their interests and can help expand their interests over time.. It enables us to not have just one Netflix product but hundreds of millions of products: one for each member profile."

Netflix uses a number of different machine learning and recommendation algorithms to drive the personalisation and search experiences. These include:

1. **Collaborative Filtering** for predicting preferences through similarities across its user base, leading to tailored content recommendations.
2. **Deep Learning** for intricate pattern recognition in user behaviour, refining the precision of personalized recommendations and features like custom artwork.
3. **Ranking Algorithms** for curating and sequencing content, maximizing personalization by highlighting the most pertinent titles.
4. **A/B Testing** as a fundamental method for continuously improving and refining personalization and search algorithms by testing new ideas and understanding their impact on user engagement.

The outcomes of these processes are additional features of the platform such as "Because you watched..", "Matched for you…" lists of movies / shows. Whilst this is beyond the initial core product offering of Netflix (which again is giving users the ability to watch what they want when they want it), it's showing users what they might want to watch based on an understanding of

their preferences. In other words, it's solving another huge problem for users, figuring out what to watch.

By Netflix recommending shows with the highest likelihood of a match, they achieve three things:

[1] Reduced latency in content discovery: When users are on the platform they either know what they want to watch or they don't. By automatically recommending shows directly on the home page, based on what they've watched previously (shown as a % match), Netflix can minimise the time between users entering the platform and watching something.

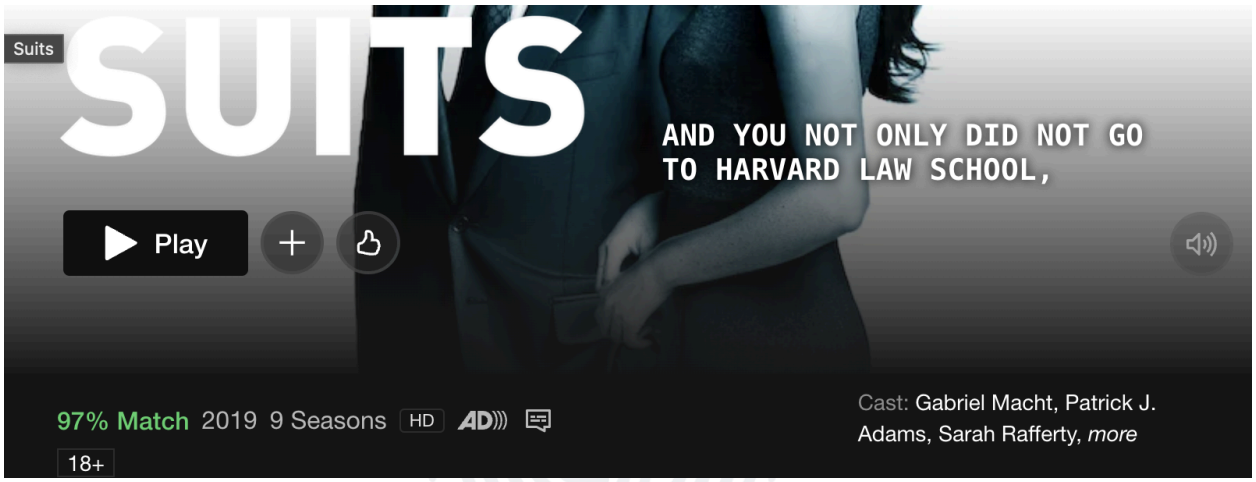


Figure 5 - The Netflix homepage comes with recommended shows that have a certain % Match based on your view history.

[2] Enhancing stickiness to the platform: When users are presented with films that align with their interests, they are more inclined to stay on the platform as it saves them the effort of searching for similar content elsewhere.

[3] Optimizing Content Catalogue: While having a diverse catalogue is important, the ability to recommend content effectively can reduce the need for an "infinitely large catalogue." By understanding and catering to user preferences, Netflix ensures that even a finite selection of content can meet a wide array of user desires, making the service feel personalized and comprehensive.

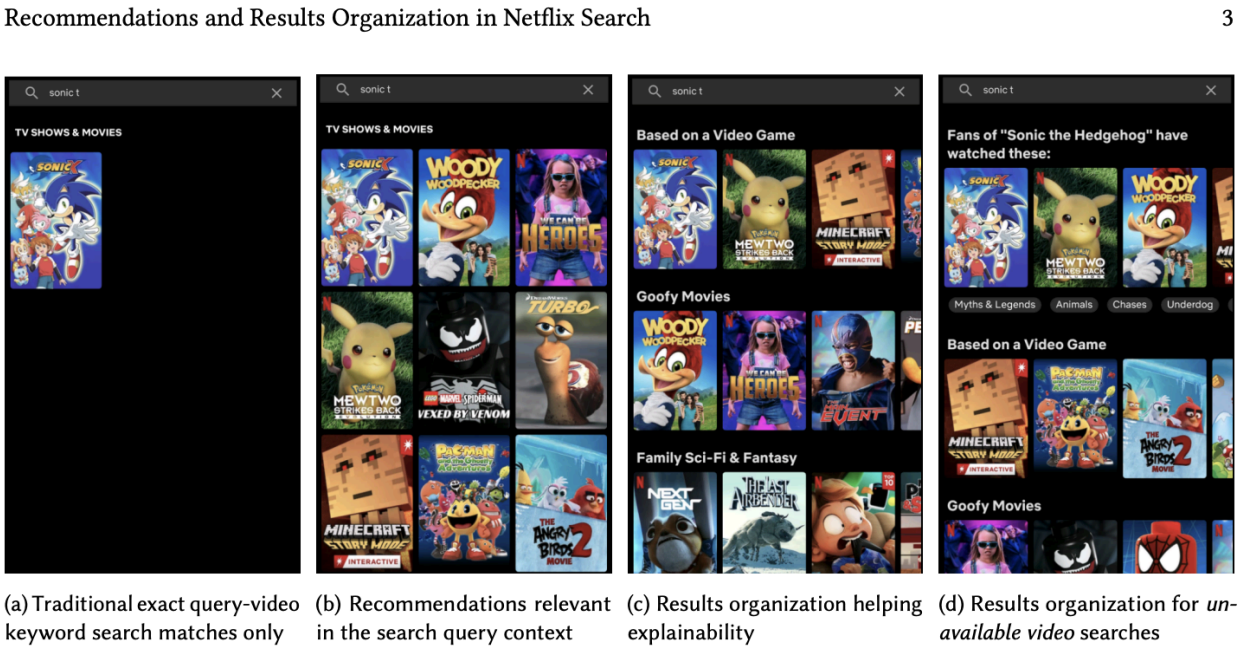
"Personalized recommendations on the Netflix Homepage are based on a user's viewing habits and the behaviour of similar users. These recommendations, organized for efficient browsing, enable users to discover the next great video to watch and enjoy without additional input or an explicit expression of their intents or goals."

Search

Netflix's search feature is powerful. It uses a combination of natural language processing, text analytics, machine learning and collaborative filtering again to build new connections between catalogue titles.

Whilst at a fundamental level, the search feature aligns with a more defensive strategy (because it directly serves the core tenet of Netflix - getting users to the shows they want to see), we wanted to mention it here because Netflix deliver value beyond simply this core necessity of search, and it links well with personalisation.

This idea is captured well here:



(a) Traditional exact query-video keyword search matches only (b) Recommendations relevant in the search query context (c) Results organization helping explainability (d) Results organization for *unavailable video* searches

Figure 6 - the evolution of search results at Netflix. (a) showing exact match results, evolving to (d) showing exact match as well as "based on.." results to showcase catalogue + inspire for further viewing.

In this image, (a) shows the results of the show that exists in the Netflix catalogue exactly related to what the user is searching for. This neither increases nor diminishes the value created by Netflix, it is neutral, as it meets user's expectations (met expectations == neutral value creation).

The following three images reflect the evolution of their search function, where now you can not only find the thing that you are looking for, but you can find similar shows which are ordered by attributes, e.g. others have also liked, based on a similar type of show, or similar category. This helps users do two things: understand that Netflix has a wide catalogue, and get

inspired/explore other shows that they may want to watch, beyond what they searched for (beyond expectations == additional value creation).

To make the importance of this more obvious we can compare it to one of Netflix's competitors, Disney Plus.

If we're searching for a specific title, e.g. "The Book of Boba Fett" (a Star Wars series), users will typically type in the most convenient keywords e.g. "Boba Fett" in this case. But the top search result is a completely different show, and you have to scroll down to find it. The first row gives you the trailer of *The Book of Boba Fett* but the real link to the series is located on the second row (8 titles away).

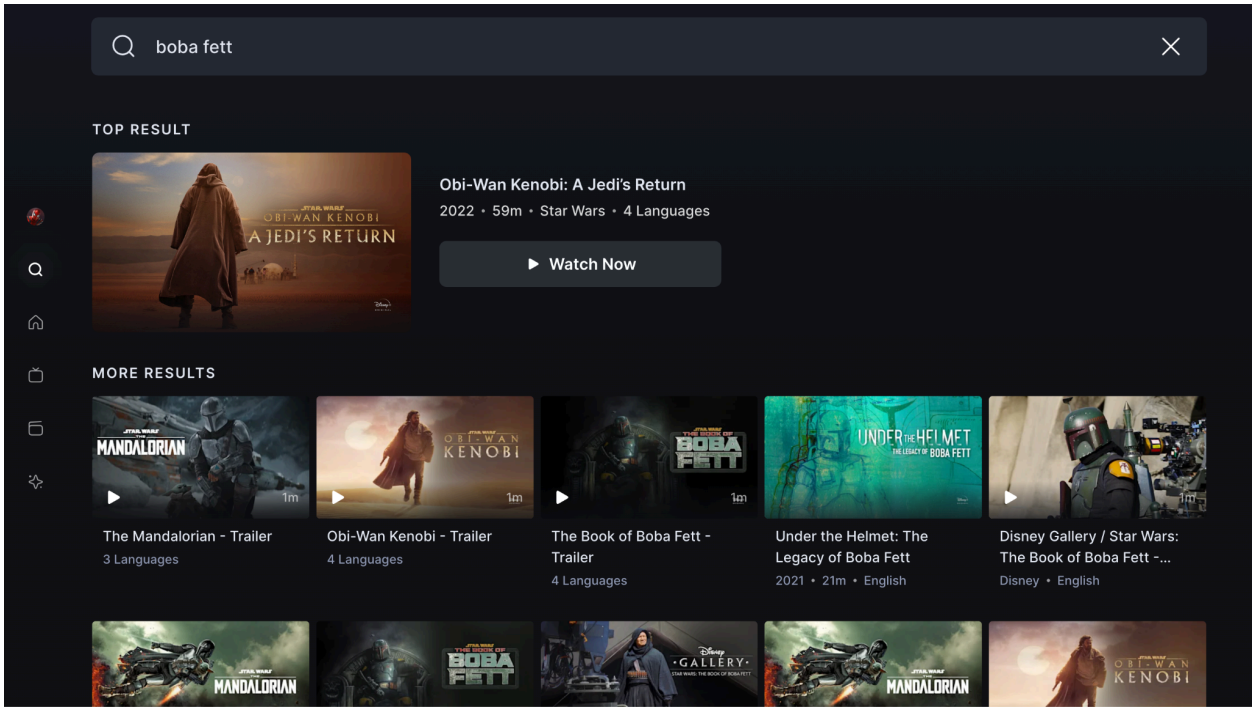


Figure 7 - Screenshot of Disney Plus when searching for "boba fett", the actual series is 8 titles away.

Even if the user goes a step further and types in "The book of boba", the top result is a completely different show again, and the user can only find it now 6 titles away.

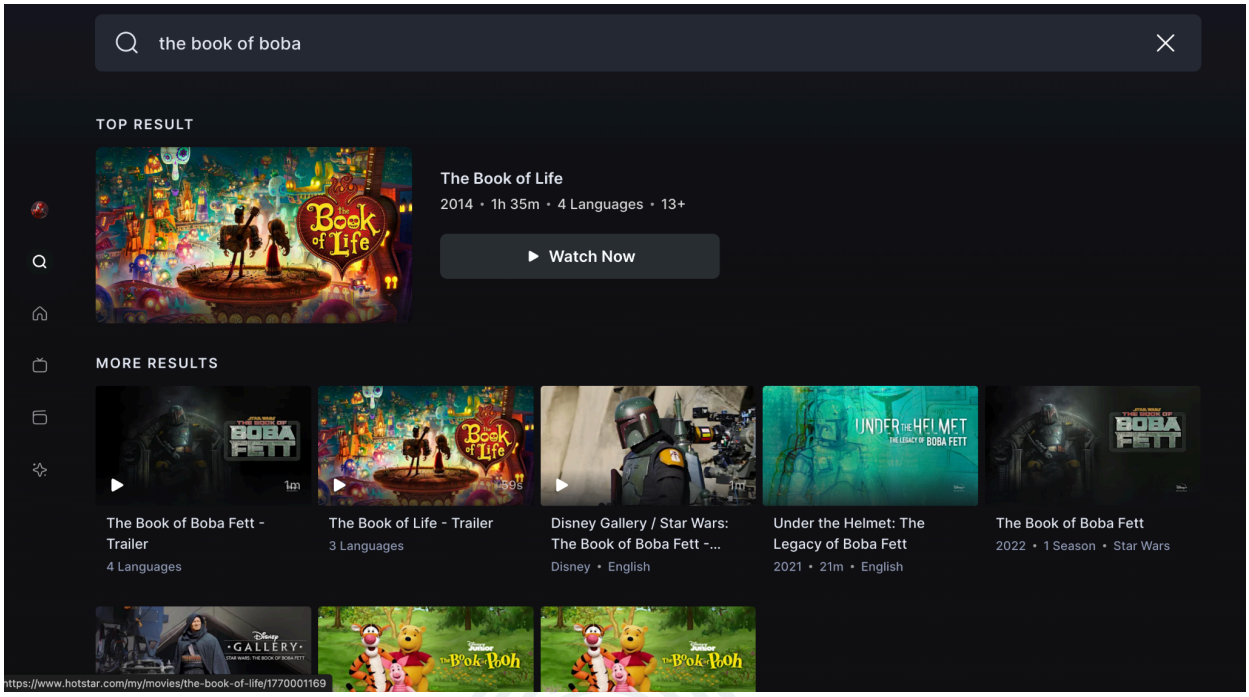


Figure 8 - Screenshot of Disney Plus when searching for "the book of boba", the actual series is still 6 titles away.

Compare this to the search of Netflix. If we search for *Naruto*, the first result is *Naruto*, with the following two results directly related shows, being sequels of *Naruto* (*Naruto Shippuden* and *Boruto*). This is an important difference in the two platforms. When a user is searching for a show, they need to know whether it is in the catalogue or not.

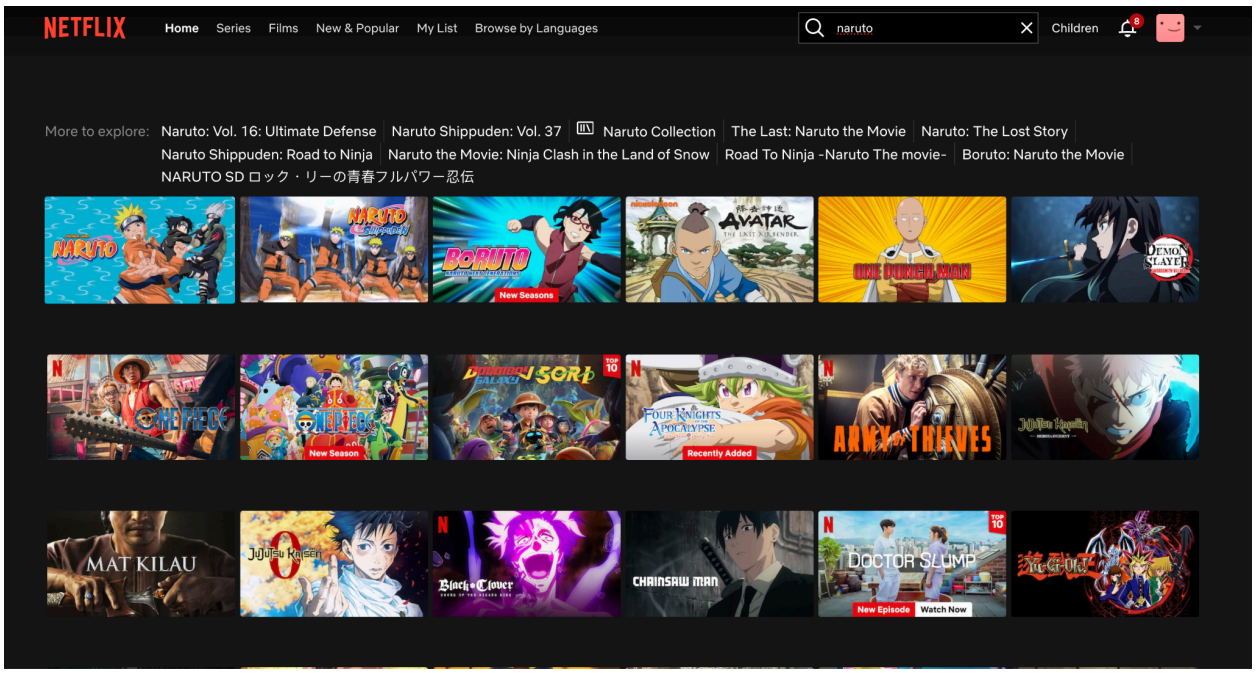
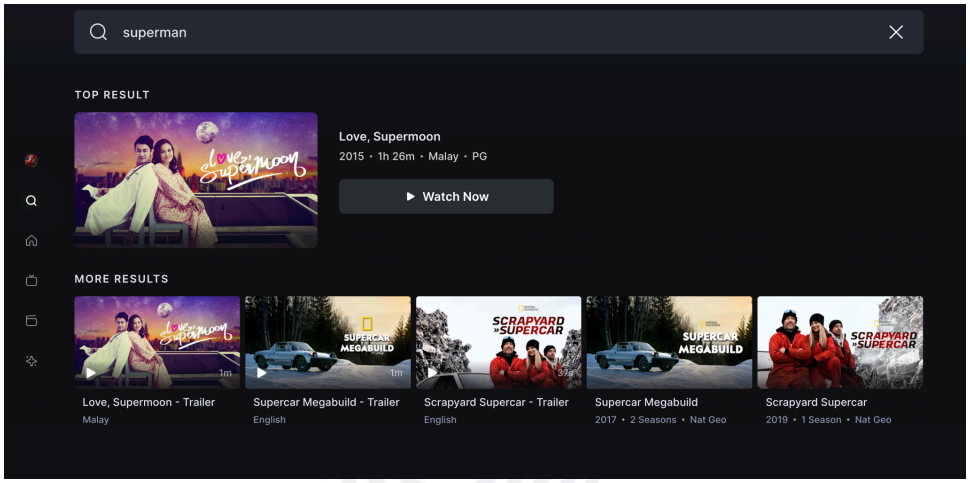


Figure 9 - Screenshot of Netflix, when user searches Naruto, it is the first title, with related titles next to it.

If a user searches for something on Disney Plus, they cannot immediately know whether it is there or not, they have to keep looking through, compared to Netflix, where if it's not the top result, they can be pretty confident it is not there. Whilst this difference may seem trivial, these product frustrations accumulate in the user's mind.

Let's take another example. There are no explicit *Superman* films in either the Disney Plus catalogue nor Netflix, yet the search results are very different and the frustrations are self-explanatory:



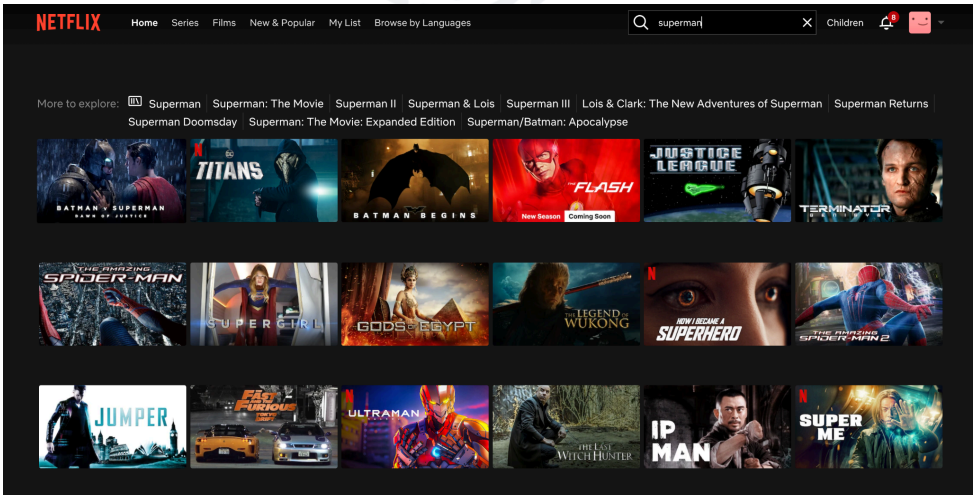


Figure 10 and 11 - A comparison between searching for "Superman" on the Disney vs Netflix platform. Whilst neither have exact titles, Disney displays shows with titles similar to the keyword searched, whilst Netflix displays shows similar to what the user is actually looking for with being explicit (i.e. superhero movies).

Whilst it may seem tedious to dive into such a niche example, these frustrations matter to the user, and because they matter to the user, they should matter to the business.

In business 101, we learn about the crucial gap between the value users perceive and the price they pay. While much emphasis is placed on finding ways to enhance this perceived value (marketing etc), the importance of reducing product frustrations is often overlooked.

These seemingly minor annoyances do accumulate in the users' minds, diminishing the overall perceived value for users. And if that value is reduced to the level of the price point, or even below, then users will not want to use your platform - something that likely contributes to what they got wrong in the Disney Plus platform:

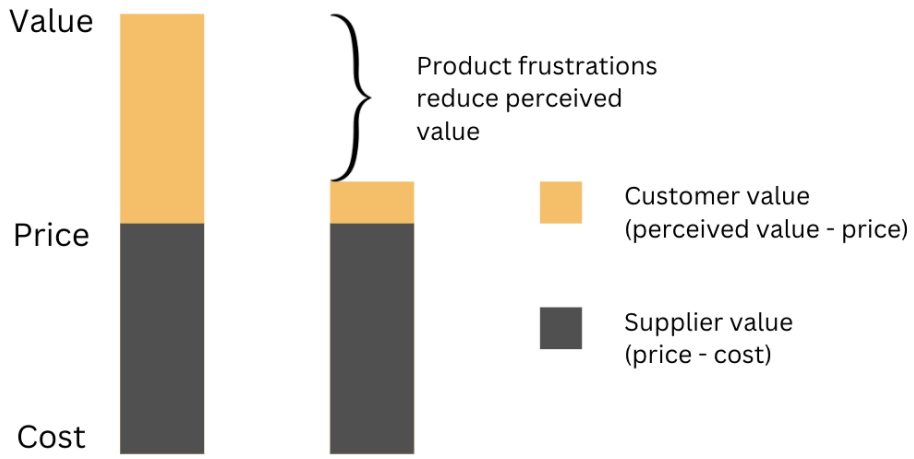


Figure 12 - A simple diagram showing how product frustrations accumulate to diminish the overall perceived value of the product by users. If it diminishes the value enough, users a likely to stop using the platform and switch to another.

Research and Development

It's also worth paying respect to Netflix's emphasis on R+D and their cross-institutional collaborations. For example, their paper on the use of variational autoencoders (a type of neural network) to improve collaborative filtering (for personalisation), was done between researchers from Netflix, MIT and Google AI.

This willingness to partner outside of just the Netflix R+D team, doesn't just underscore the strategic advantage of leveraging collective expertise across industries and academia, but something deeper. It represents an attitude and culture at Netflix, where instead of being obsessed with trying to *increase their slice of the pie* (competitor-focused), they are obsessed with *increasing the size of the whole pie* (user-focused), and in doing so continuously redefine the landscape of digital streaming and entertainment.

OTSOG©