

LEASEWEB CUSTOMER CASE

YUBO

Yubo is a social media app designed for teenagers and young adults to meet new people via live streaming and messaging.



It's a mindset that's deep in the culture of Yubo: build things that scale.

- Arthur Patora, Co-founder & CTO, Yubo

The Customer

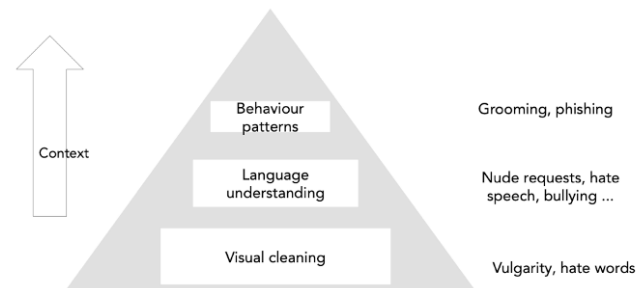
The team behind social media app Yubo is paving the way for smartly built apps that pack a lot for their punch. With a lean team supporting more than 50 millions (and counting) users on a combination of Leaseweb dedicated servers and custom-built AI systems, Yubo stays laser-focused on creating a safe and scalable application.

The Challenge

Yubo needed to ensure the safety of their users by moderating content - mainly text, images, and live streams (known as 'lives'). "Our mission is to connect people and allow them to interact with people they don't know in real time," says Arthur Patora, Yubo Co-founder & CTO. "To serve this mission, we need to make the app as safe as possible. Everything we do is to empower our users."

To do this, Yubo created tons of in-app features allowing users to moderate their own profiles and 'lives'. For everything else, the team turned to AI.

Yubo uses AI primarily for facial recognition, ID verification, and text analysis. As the app grew, user language evolved too - and began to take on a life of its own. "We used a third-party API in the beginning, but it was not matching all of our needs," says Patora. "We had too many algorithm change requests per iteration, plus most of our content is short-form — something that text moderation tools can't pick up on as they are mainly



yubo

15 million
messages sent on the
app daily

1.55 million
active learning samples
annotated monthly

Customer Summary:
Yubo is a social media app designed for teenagers and young adults to meet new people via live streaming and messaging.

Industry:
Social media, streaming

Founded:
2015

Headquarters:
Paris, France

Size:
11 - 50

Products in this case study:
• Dedicated server
• Private network

“If the moderation is down,
then everything goes down.
It’s a nightmare.”

- Arthur Patora, Yubo Co-founder & CTO

trained on articles.” Patching up their text moderation problems was proving costly and resource intensive. “We even started going into API interfaces to type the wrong words we found on the app that our human moderators saw all day,” says Patora. It wasn’t sustainable or scalable. “That’s when we realized we needed to build our own system.”

More than just text moderation

But just how unique is the text in Yubo? Alexis Benichoux, Machine Learning Specialist at Yubo, breaks it down:

- It’s hard to spot which language is being spoken because many users mix several at once. For example, a user may say something acceptable in Swedish but then be flagged as being toxic in English due to a different meaning of the word in the English language.
- Most messages are very short, and many short slang words may be common in one language but mean something completely different in another.
- Some languages are very similar, such as Spanish and Portuguese.
- There is Internet slang and even “Yubo language” that is specific to the app only.
- Many messages use emojis — one message can just be three separate letters with an emoji.
- Intentional misspellings of profane or prohibited content are common and also need to be accounted for.

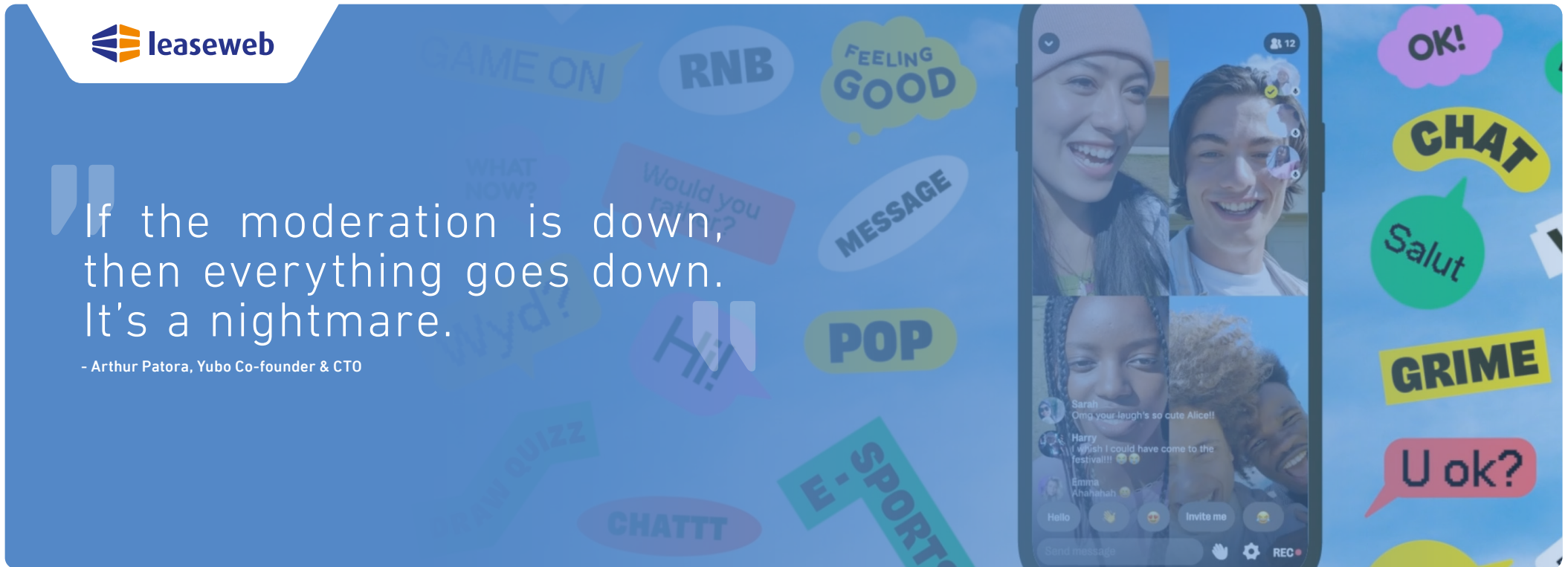
The Solution

Yubo decided to create their own moderation platform and beelined for a dedicated server solution at Leaseweb. They had experimented with other third-party cloud solutions in the past but ended up transitioning to dedicated server solutions due to their lower costs and high scalability potential. “We were originally built on Google Cloud,” says Patora, “but it quickly became way too expensive.” When they switched from Google to Leaseweb, Yubo saved over 80% of what their infrastructure costs would have been. This money could then be invested back into their employees, the new moderation platform, and the app.

With bare metal as the basis, Benichoux and his team (dubbed ‘Team Profanity’) built out the AI moderation platform, defining and filtering the data into something processable and workable.

System architecture

The models are based on two types of content: internet content (containing mostly high-volume slang) and internet trolls. For this, fast APIs can be deployed in production to moderate content. The second type of content is toxic users & imposters, which requires more machine learning and knowledge to filter properly. While this content is lower volume, it is much more potentially damaging for the app and users. Low volume content is treated with deeper algorithms in asynchronous time, whereas high volume needs fast responding APIs. Once the content type is



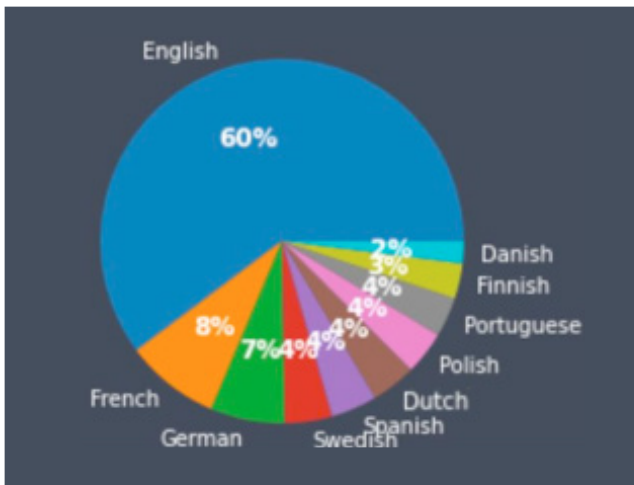
Text moderation at Yubo is not really language processing — it's something very specific.

- Alexis Benichoux, Yubo Machine Learning Specialist

identified, it enters a data flow consisting of different models. Each language, category, and feature have different constraints. There are about 120 models in production. These are based mainly on (CBOW+fasttext and NBSVM). For example, a chat-like message will filter through at least three models. A typical pipeline involves:

1. Stemming, lemmatization, deobfuscation
δοσ γαλλ ωαηηα νε φιαηηδς -> (do, you, want, be, friend)
2. Language detection

Language Repartition

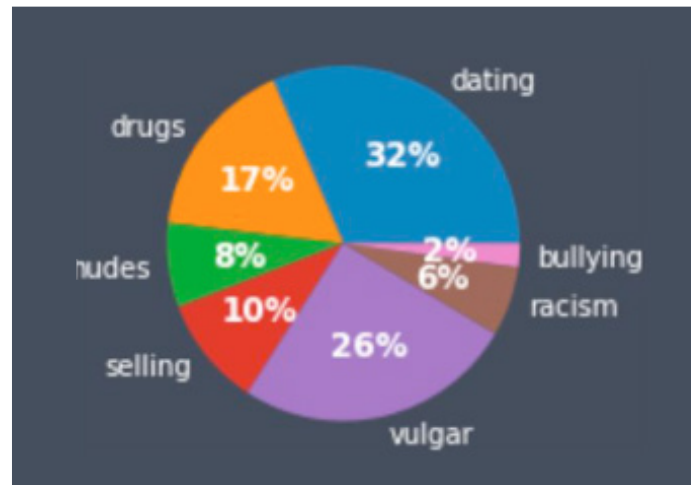


3. Personal information detection
4. Profanity filtering

Data flow

- Model training
 Yubo data is used to constantly retrain models. This is done offline, and checkpoints are sent to data buckets to be served by frontend APIs.
- Data: the team starts with a clean data set, which can be entirely outside of production. No outside data is used

Categories



Key Requirements

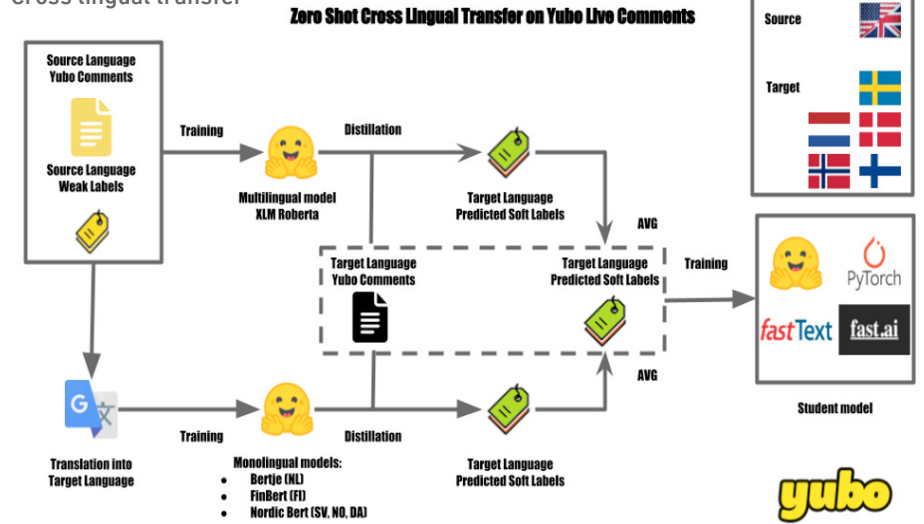
- **Scalability:**
 Yubo's userbase grows fast, so they need infrastructure that can grow with it.
- **Secure network:**
 Yubo conversations may contain private content, making it essential to have private networking in the infrastructure with support for end-to-end data encryption on top.

Our Solution

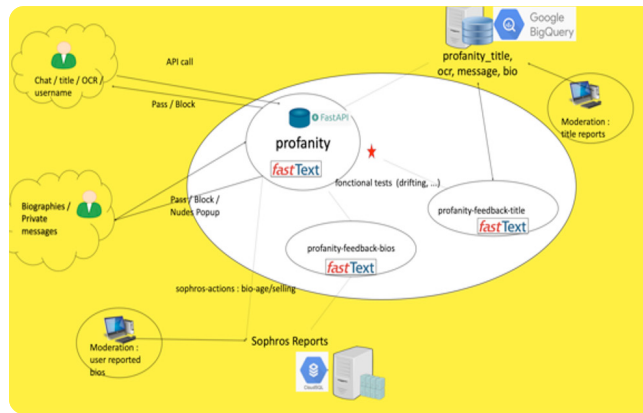
- Dedicated high performance servers
- Private networking enabling interconnection of servers deployed in Leaseweb data centers for increased scalability and security purposes
- High performance CPUs (E5-2650v2) for analytics and standard CPUs (E3-1240) for proxy servers
- SSD storage for optimal speed to retrieve, treat and exchange data
- Services in the United States, Netherlands, and Australia



Cross lingual transfer



- everything comes from Yubo.
- Annotation: there is an app user-based annotation as well as annotation campaigns.
- Modelling: different models exist for various features (such as live chat, biography edition, image uploads, etc.).
- Production: the model is categorized and put into appropriate production environments
 - High volume content: fast answer needed (messaging, etc.) - APIs are deployed, computation is done on GPUs.
 - Low volume content: more processing time is needed (investigation into a toxic user, more context needed, etc.) - use Redis Queues, learning feedback trained and deployed. Use a smart supervised learning model.



Supervised to active learning

"The front part of the architecture is trained models," says Benichoux, "and then there is a big backend part. The team is mainly training and deploying models, so it's supervised learning, but the big picture is that we are trying to put active learning into supervised learning. So, we're using human moderation - anything we can get, even if it's very slow to trying and modify it - but that's how they look, and we get human reports. Some looks can even take weeks. This is how we've achieved slow but steady improvement in deploying our supervised learning algorithm."

The Outcome

Yubo's text moderation architecture is only two years in the making and is already one of the most intelligent systems out there. Among other initiatives, Yubo has partnered with National Center for Missing & Exploited Children (NCMEC), an American NGO dedicated to the search for missing children, by sharing data with the organization if there is any suspected illegal activity involving minors.

In addition to the technical resources guaranteeing users' safety, Yubo is also supported by a Safety board made up of the best international experts. The Board (with experts coming from Thorn and Interpol, for example) meets several times a year to go through the program of product features and reviews the safeguards in place.

Every day, the algorithms become more intelligent, and Yubo's user base (and infrastructure) grows. The team keeps things stable and straightforward while safely connecting thousands of new users every day — proving that safety does not need to be sacrificed for scalability.

The Future

Yubo has the ability to scale exponentially — now, the next step is to recruit talent fast enough to support the growing company. And with horizontally scalable architecture containing innovative AI systems powered by Leaseweb dedicated servers, Yubo's going to need a lot more talent.

Next step? World domination.



Get in Touch