



HAL
open science

CAPTCH'YOU

Melike Aydinlilar, Alaaeddine Chaoub, Kelvin Han, Florian Marchal

► **To cite this version:**

Melike Aydinlilar, Alaaeddine Chaoub, Kelvin Han, Florian Marchal. CAPTCH'YOU. THINK BEFORE LOADING, , 2024, 978-2-9591975-0-5. hal-04518973

HAL Id: hal-04518973

<https://hal.univ-lorraine.fr/hal-04518973>

Submitted on 24 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



NE VOUS EN FAITES PAS, ÇA VA MAL SE PASSER

THINK BEFORE LOADING



Melike Aydınllılar
Alaaeddine Chaoub
Kelvin Han
Florian Marchal-Bornert



*La menace est grande pour le système,
les autorités doivent agir rapidement et brutalement
afin que le système garde sa stabilité.*



Traduction française par les autrices et auteurs.

Ethics in Computer Science:

Write your Dystopia

15 November 2022

Catch'you



From:

Melike Aydinlilar <melike.aydinlilar@inria.fr>

Alaaeddine Chaoub <alaaeddine.chaoub@loria.fr>

Kelvin Han <huiyuan.han@loria.fr>

Florian Marchal-Bornert <florian.marchal@loria.fr>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Chapter 1

It's 7am and Laura wakes up. After her morning routine, and while having her breakfast, she opens Captch'you™ -- the citizen's app -- to check what the other members of her committee decided overnight. Like everyone, she helps to keep the city's society thriving by judging whether the activities of others are acceptable or not.

Laura cannot actually see the daily lives of others, but every adult in the city is part of a system that surveys and monitors the city itself. Twice a day, the system assigns committees a set of short video clips, with potentially problematic behavior detected that day for them to classify. This is part of life in the city: everyone has a moral obligation to review and assess the clips their committee gets each day.

After fulfilling her basic needs and moral duty for the day, Laura leaves for work. While waiting downstairs for the pod to the bus station, she remembers...

The light-switch! It's still on.

Afraid of the warning that the system will generate -- for wasting energy -- and that it will flag her action (or inaction) for committee judgement, she runs back to the apartment and switches off the lights.

Thank goodness!

It is 8.30am and Laura arrives at work. Her job is special -- as a moderator within the system, she is responsible for making the final decision on Captch'you™ cases that have been judged by the committees as reflecting inadequate social behavior. Because of that, she is granted access to more than just the short clip that others receive on the citizen's app, and she can recover fuller clips and personnel dossiers for each case if she wishes to.

While going through her pile of cases for the morning, Laura receives one where she has to decide whether a woman can keep her child or not. There have been nine other committee decisions so far against the woman's behavior towards her child. In the clip sent for classification this time, she can see the mother yelling at the child. So, she will support their decision, and approve the outcome produced by the system -- the child should be taken away to a more caring family.

Ten times... It was not that hard to decide, this one.

Laura pauses for a moment though. She can see that this mother really looks like her. Maybe it is the woman's hair or maybe it is her age, but something causes Laura to relate to the woman. However, carried away by other cases waiting, and thoughts about the dinner she planned with her parents later that night, she does not spend time thinking further and taps on the 'Bad' button on the screen.

After working her day as a community servant, Laura heads back to her apartment to start preparing dinner.

The doorbell rings. Anxiety suddenly jolts Laura.

Is everything ready? The carrots are in the stew. The sauce is not too hot this time. The dining room is clean. A few seconds of waiting before answering is usually judged normal by the system, but I mustn't panic and jump to the door, otherwise the system could...

Her parents enter the apartment together as they usually do. Her mother hugs her, her father kisses her cheek and all three gather around the dining table. Laura starts serving dinner. As she settles into her seat, her mother looks at her hair with a forced smile. Then it comes:

- You didn't take time to be pretty, right?

Despite knowing well that this could happen, Laura still feels attacked. A stab in the back precisely. She breathes deeply to calm herself down.

No retorts, I must behave like a good citizen.

She let the remark slide and leaves the kitchen. She takes the time to fix her hair, but what her mother said about it reminded her of the woman from the morning.

Am I that unkempt? Was the woman that unkempt after all?

When she comes back she offers wine to her parents and sits down to finish her cold course. After dessert, she makes tea for everyone, and they start to talk about little things. Then the woman comes back to her mind.

- You know, this morning I had to judge a woman. The crowd stated that she made a lot of tiny little mistakes. I'm sure the system is right, but I felt like I was missing something? I don't feel good about it.

Her father simply nods.

- You did what you had to do.

- Yes, but she kinda looked like me. It could happen to me one day, you know.

Her mother frowns.

- No. You're stronger than that.

Laura shifts uncomfortably.

- They will take her child off her you know.

Sadness falls on her parents faces. A silence, probably only two seconds from being detected by the system as being too long, hangs over the air. Dad starts to say something, but her mother hit him in the side with her elbow. Mother answers sharply.

- Still, you did the right thing. You did your job. Be proud of it.

The conversation continues onto more “harmless” subjects like Laura's single status, the weather and her parents’ preoccupations. The night could have just ended that way -- after all her parents acted normally when they said goodbye to her. But after the door shuts, she watches them from the window of the apartment.

Occupational hazard of mine, I can't stop monitoring...

Down there, near the building entrance, they seem to be arguing loudly.

Are those tears on mother's cheeks?

Laura flees this vision, baffled. She finds herself having to make more effort than usual to end her day like a good citizen. The memory of the woman constantly comes back in her head.

Who is she? Do my parents know her in some way? Why does she resemble me in some ways? If it is the case, how?

Not knowing it then, but what Laura understood as reality slowly unravels from this point.

Chapter 2

The next day, she goes to work still thinking about the dinner with her parents. The woman from yesterday is still bothering her. After taking her coffee, carefully without making a mess in the break room, Laura sits at her desk, and starts her day.

Some litterers. Nothing new. Some traffic rule violations too.

She is relieved that at least there is not much violence on her screens this morning.

At noon, she grabs a sandwich, and somehow, she finds herself going back to the clip from yesterday. She opens the case dossier and gets the nine earlier clips against the woman together with their extended versions too.

Here we go.

She wants to explore her bad feeling around this case.

She tries to watch a little bit more around the timestamp of yesterday’s "incident" clip. She starts with the yelling scene.

Classic.

She has seen thousands of parents losing their self-control before.

There should be mandatory anger management classes for everyone. At least there is no hitting in this clip.

This time however, she notices some concern in the facial expression of the woman. Rewind, rewind, play. It is a kitchen scene, there is a hot pot on the stove, the handle is turned outwards.

How careless!

The toddler starts reaching towards the handle. Fast-forwards 15 seconds: The mother yells to stop him in time.

Another clip. The toddler is crying in a playpen, yet the woman is moving out of the frame.

Doesn't she know that this is not good for child development?

Fast-forwards 10 seconds: The mother is back soothing the child. It's a lovely scene, the mother brought the small toy elephant that appears to have been pushed out of the playing area back to the child.

The feeling sinks in Laura's stomach, making her more uneasy.

Why were these clips flagged in the first place?

She goes over the other clips nervously. It is almost at the end of her lunch hour. She is not supposed to spend this much time on one single case, and she will have other cases to evaluate.

It never ends.

Rewind. Fast-forward. Over and over again.

It looks like all the woman's cases have some benevolent explanation. There has to be something she is missing. Laura checks the background, the living room, kitchen, but there is no sign of mold on the ceilings.

Maybe lead paint?

No, the house looks new enough. There is also a healthy number of children's books on the bookcase. There are even child-locks on the drawers.

What is it that the system keeps flagging this case?

She returns to her duty after lunch, works hard and tries to keep the case out of her mind.

Finally, it is 6pm and time to go home. Her fellow community servants are leaving one by one. She thinks to pack up, and maybe take the case to her supervisors later. But she ends up glued to her desk, trying to find something in the clips. She checks the picture frames on the walls and dressers in the woman's home.

She finds a young man in most of the pictures.

Funny, I've not seen him in any of these ten case clips so far.

Then she realizes the first clip surprisingly started only a few months ago.

Usually the problematic cases are spread over longer periods, starting with small aggressions and petty crimes. She watches the earliest video. This one shows the mother crying with the child on her lap.

Is it emotional dysregulation?

Rewind, rewind. There are more people in the apartment.

Can it be? ... Oh my God! She lost her partner around the time the clips had started to pop up in the system.

Laura has to leave. Staying too late could flag her. She goes home with a thousand thoughts in her mind.

Could the system be penalizing her just for that? Is this case an anomaly? How many decisions have I made with clips that are similar to this case?

She spends the next weeks scrutinizing every single clip she encounters in her caseload, every now and then she starts getting the same uneasy feeling.

Something is just not right.

Chapter 3

It is two months from the day she validated the decision to remove the woman's child. Laura cannot stop thinking of the sorrow the woman must be going through, and the distress the toddler must be in, to be in a strange home without his mother.

Laura goes into the office determined to act.

She has realized that the city system is not being used to help citizens judge their fellow citizens' actions. All the outcomes produced by Captch'you™, including taking children away, have already been determined by its automated forecasting model beforehand.

She investigated. The clips that are flagged for committees to review are only for show, the prediction model has been trained to suggest cuts to the video footage so the clip shown to the committee will bring about the outcome the model has already decided.

In another cruel twist, it turns out also that the system itself is actually judging everyone based on the ratings they are giving. Nothing Laura or her fellow citizens thought they were doing actually made a difference. Worst yet they were, and they still are, feeding data back to the model to rate themselves.

She logs into the system. She knows that she must do something. A large pile of new moderation decisions is awaiting her today. She starts feeling sick in the stomach again.

How many parents are going to face the consequence of a decision today that will drastically change their lives? Will there be a single parent going to spend a sleepless night...? Their child in someone's home just because a pot on the stove and their child was unattended together in the kitchen for five seconds, or did not reach out to soothe their child within ten seconds?

The idea for a small act of rebellion swelled from within her. She begins clicking 'Good' on the first case decision without checking. She starts to feel a little better.

Then she does it again, and again, on every single decision that she has to make. A feeling of satisfaction fills her, but it does not feel enough.

The next thing she does will change everything though.

Epilogue

Laura steps out of City Hall. A police car is already waiting for her on the plaza. She is handcuffed and shoved into the back seat. The officers say nothing and only give her a dirty look. It is enough to let her know how they feel about what she did.

This time, fear suddenly jolts her. She worries about what will come next, she has seen terrible things happen for lesser “crimes” against the system, let alone something committed by a person of her position. The threat to the system is great, the authorities have to act quickly and brutally so the system is not destabilized.

It is going to be a sleepless night for her, or worse. But she also strangely feels a sense of peace and calm for believing that she did the right thing.

As the autopilot on the police car kicks in and they are driven onto the highway out of the city, one of the officers instinctively pulls up Captch'you™ and was about to start rating people on the screen, however, all the panels are only showing a single clip of a masked figure saying: *"Wake up, the AI fools you"*. The clip is short and goes on repeat just like the usual ones, but this time it ends with something more before repeating: a URL button flashes on the screen.

Laura did it.

Her bomb dropped where it was meant to – all the case files and extended clips for the millions of committee judgements that were made over the last two years are at that URL, starting with the ones involving child protection rehoming.

It was not that hard to decide, this one.

A smile forms on Laura's face, but she pauses mid-smile.

What if everyone else thought the same as the police officers?

At least Dad might say: “You did what you had to do.”, but a video from the morning's pile still bugs her.

Did the father who fed his son expired apple sauce really not know he was doing that, or is he really a negligent father?

.....