

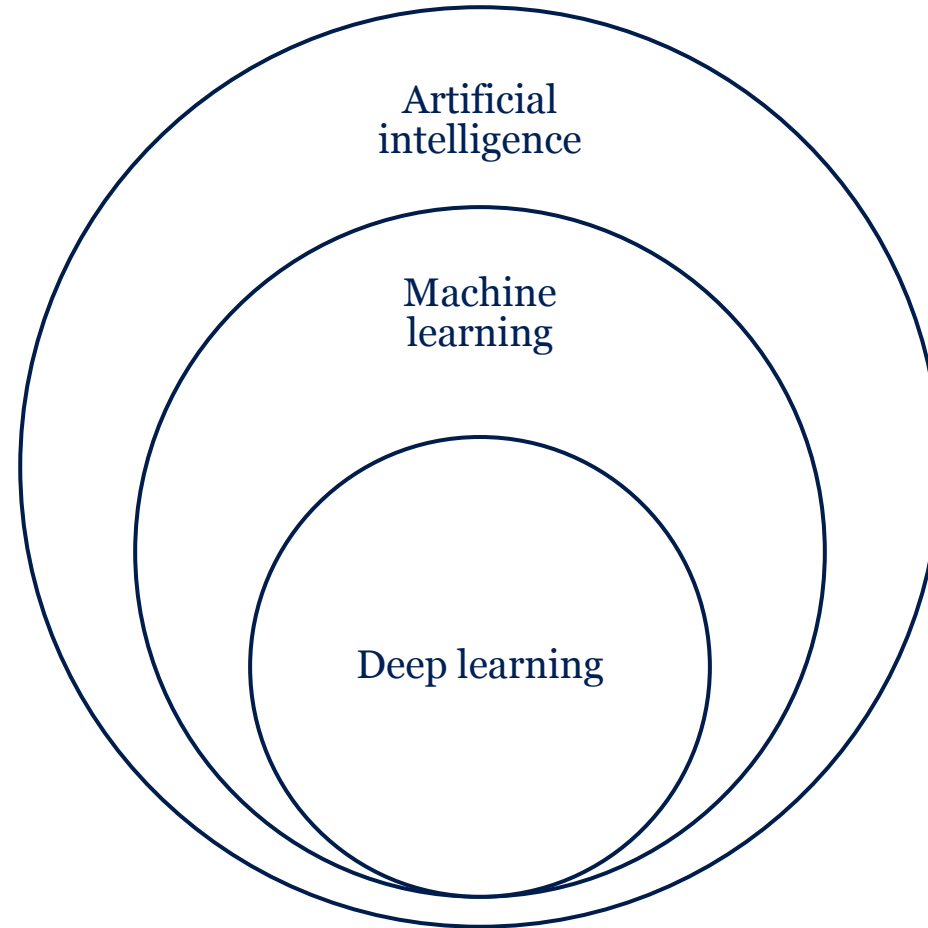
# Using machine learning to boost internet and DNS security

Thymen Wabeke | EDNS

4<sup>th</sup> of October 2021



# Machine learning in perspective



# Research agenda

- Apply ML to increase security of the Internet and DNS
- Approach: explore and integrate promising algorithms, papers and tools
  - Innovating *with* ML, not innovation *of* ML
- Target group: DNS actors (registries, registrars and DNS operators)

# Research topics



RQ1: How can we get even better at proactive abuse detection?



RQ2: How can we train shared abuse models without exchanging data?




RQ3: How can we use ML to improve our anycast infrastructure monitoring and management?



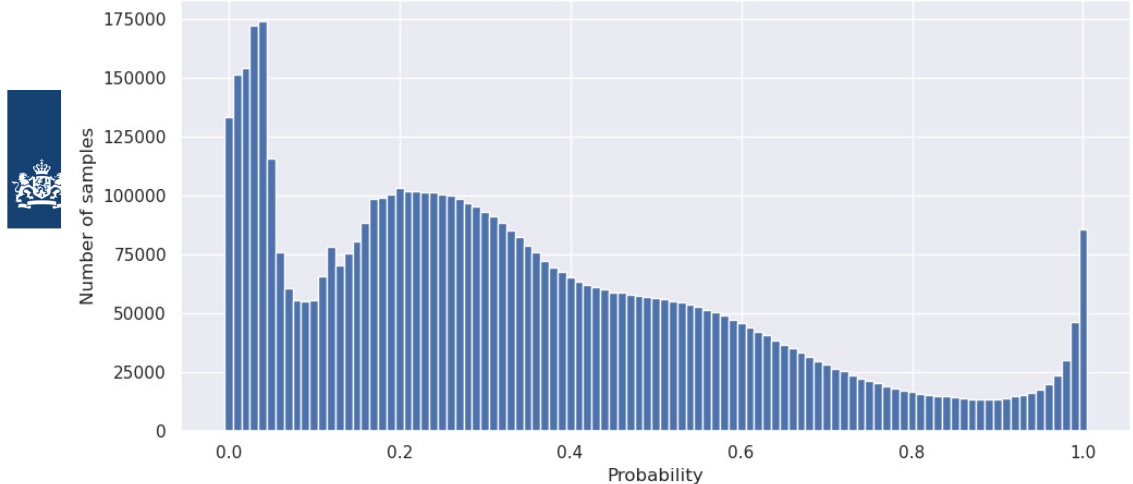
# Applying ML in a responsible way

- Human-in-the-loop
- Simple and interpretable models
- Collaborate and publish
- Monitor performance

**Radboud University**  **REALTIME REGISTER**

Counterfighting Counterfeit: detecting and taking down fraudulent webshops at a ccTLD

Response Distribution Chart




UNIVERSITY OF TWENTE

for a large array of goods, such as pharmaceuticals, are among the top 100 from the consumer perspective.

ICCS VISA MasterCard INTERNATIONAL CARD SERVICES

Thursday 10 June 2021  
Article by: Thijs van den Hout, Thymen Wabeke, Cristian Hesselman

The [original blog](#) is in Dutch. This is the English translation.



# Remainder of presentation



A photograph of two mannequins in a retail setting. The mannequin on the left is wearing a dark patterned shirt and a hat. The mannequin on the right is wearing a red top and a red headscarf. Three circular callouts are overlaid on the image: a red circle with '50%', a green circle with '30%', and a yellow circle with '20%'. A dark blue banner at the bottom contains the text 'FaDe'.

50%

30%

20%

FaDe



A close-up photograph of a large pile of various beer bottle caps from different brands, including Heineken, Tsingtao, and others. A dark blue banner at the bottom contains the text 'LogoMotive'.

LogoMotive

**HOLLISTER** Dames Heren Inloggen Register (0) Omschrijving

<p>★★★★★</p> <p>Hollister Undergoed Heren Lage Taille Multipack Grijs Groen Camo Zwart 11859-FNX</p> <p>15 Kleur <b>BROEK &amp; KORTE BROEK</b></p> <p><del>€30.60</del> <b>€22.31</b></p>	<p>★★★★★</p> <p>Hollister T Shirt Heren Logo Graphic Korte Mouwen Donkerblauw 21170-HLT</p> <p>15 Kleur <b>TOPS</b></p> <p><del>€30.70</del> <b>€22.38</b></p>	<p>★★★★★</p> <p>Hollister Jas Dames All-weather Stretch Sherpa-lined Zwart 45801-GXL</p> <p>1 Kleur <b>JASSEN</b></p> <p><del>€98.35</del> <b>€69.73</b></p>	<p>★★★★★</p> <p>Hollister Joggingbroek Heren Advanced Stretch Cargo Twill Olijfgroen 97393-SXN</p> <p>4 Kleur <b>BROEK &amp; KORTE BROEK</b></p> <p><del>€50.11</del> <b>€35.98</b></p>	<p>★★★★★</p> <p>Hollister Blouses Dames Fluweel Off-the-shoulder Goud 49289-JQI</p> <p>2 Kleur <b>TOPS</b></p> <p><del>€30.60</del> <b>€22.31</b></p>
<p>★★★★★</p>	<p>★★★★★</p>	<p>★★★★★</p>	<p>★★★★★</p>	<p>★★★★★</p>

# SIDN's interest

- Consumer losses
- Trust in Internet may decrease

## Perfect vantage point:

- List of *all* .nl -domains
- Passive and active measurements

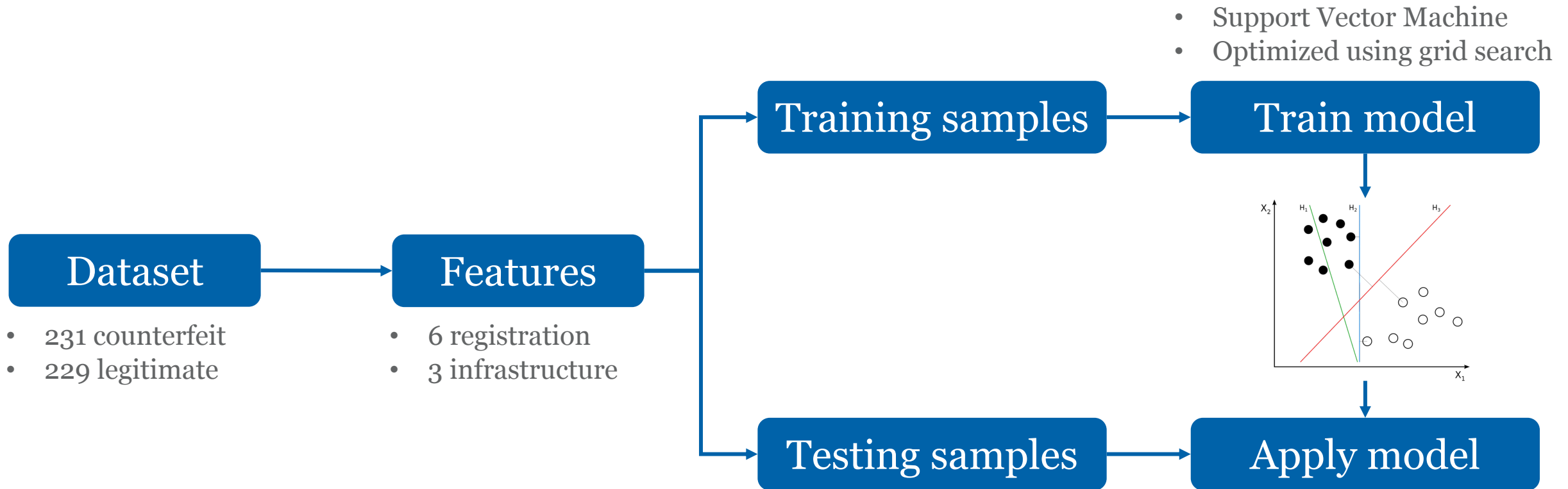




# Main results

- Detected thousands since 2016
- Protected users from being scammed
- PAM2020 paper:
  - BrandCounter (2018 Q1-2)
  - FaDe (2019 Q1)

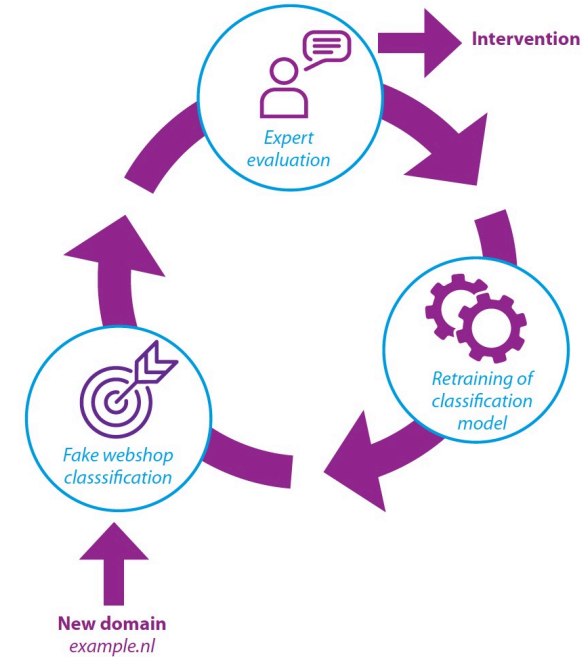




Samples	Precision	Recall
Train (cross-validation)	0.98	0.97
Test	1.0	1.0

# Lessons learned

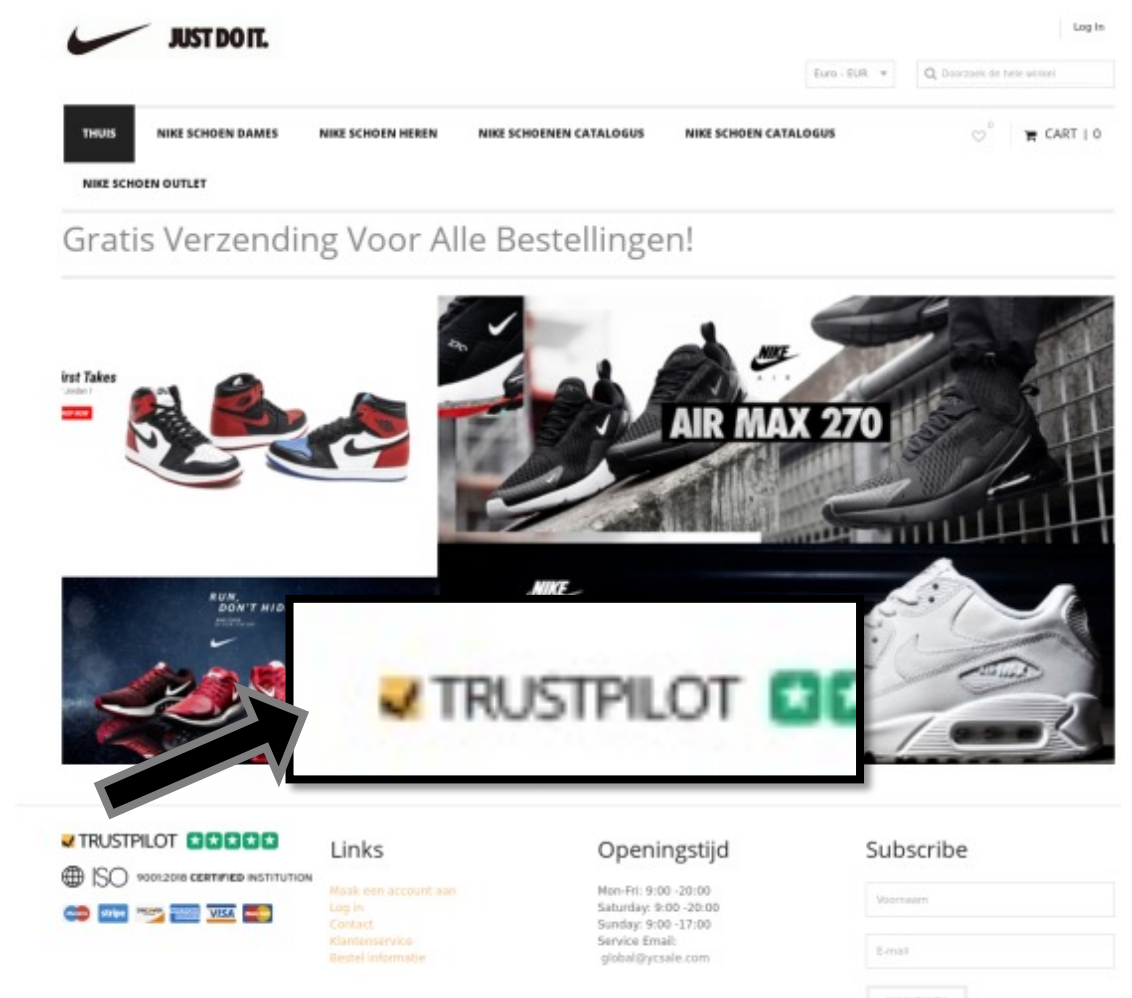
- Registrar and ICS collaboration was key
- Detectors are simple yet effective
  - Registries have perfect vantage point
  - Suggests little pressure
- It's an ever-going whack-a-mole game
  - Monitor features and evaluate model regularly
  - Fewer takedowns = fewer scams?



Year	Taken down
2018	~12,000
2019	4,340
2020	481

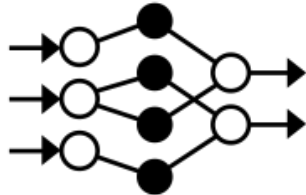
*Number of counterfeit webshops taken down*

# Malicious websites use well-known organizations' logos





# Help analyst find suspect websites using logo detection



## 1. Crawl and take screenshots

- Should be efficient: skip duplicates
- Using Selenium

## 2. Object detection algorithm

- Should be flexible: no manual labeling
- Based on YOLOv5

## 3. Annotation dashboard

- Should be easy-to-use and generic

## Filter

Label

All



Status

All



Filter

## Bulk update

Label

Select



Status

Select



Update

## Logo's found

Show  entries  Select AllSearch: 

Domain name	↑↓ Screenshot date	↑↓ Registrar	↑↓ Registrant	↑↓ Registered on	↑↓ Label	↑↓ Status	↑↓
sidn.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	1999-11-18	-	Open	<a href="#">Annotate</a>
domainregistry.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	2001-03-27	-	Open	<a href="#">Annotate</a>
dnsops.nl	2021-08-12 11:11	Stichting Interne...	Stichting Interne...	2007-09-26	-	Open	<a href="#">Annotate</a>
dnssec.nl	2021-08-12 01:11	Stichting Interne...	Stichting Interne...	2001-03-26	-	Open	<a href="#">Annotate</a>
6miljoen.nl	2021-08-11 22:08	Stichting Interne...	Stichting Interne...	2020-06-12	-	Open	<a href="#">Annotate</a>
sidnlabs.nl	2021-08-11 10:18	Stichting Interne...	Stichting Interne...	2010-05-10	-	Open	<a href="#">Annotate</a>
abuse204.nl	2021-08-11 01:39	Stichting Interne...	Stichting Interne...	2014-08-11	Correct use	Afgehandeld	<a href="#">Annotate</a>
otic.nl	2021-07-16 09:47	Stichting Interne...	Stichting Interne...	2010-04-01	Correct use	Afgehandeld	<a href="#">Annotate</a>

Logo found on [sidn.nl](https://sidn.nl)**Screenshot date** 12-08-2021 11:11**Page also found on** [dnsops.nl](https://dnsops.nl), [schijtbakkes.nl](https://schijtbakkes.nl), [domainregistry.nl](https://domainregistry.nl), [vdstbv.nl](https://vdstbv.nl)**Registrant** Stichting Internet Domeinregistratie Nederland**Registrar** Stichting Internet Domeinregistratie Nederland 2**Registration date** 18-11-1999 00:00**Screenshots**

Zorgeloos online

Producten ▾

Over SIDN ▾

SIDN Labs ▾

SIDN fonds

Actueel ▾

Contact EN

**Comment**

Comment...

Clear label

Previous

**Label**

- Correct use
- Incorrect use
- Geen logo

**Status**

- Open
- In behandeling
- Afgehandeld

Save and update all  
related domains

Save and next

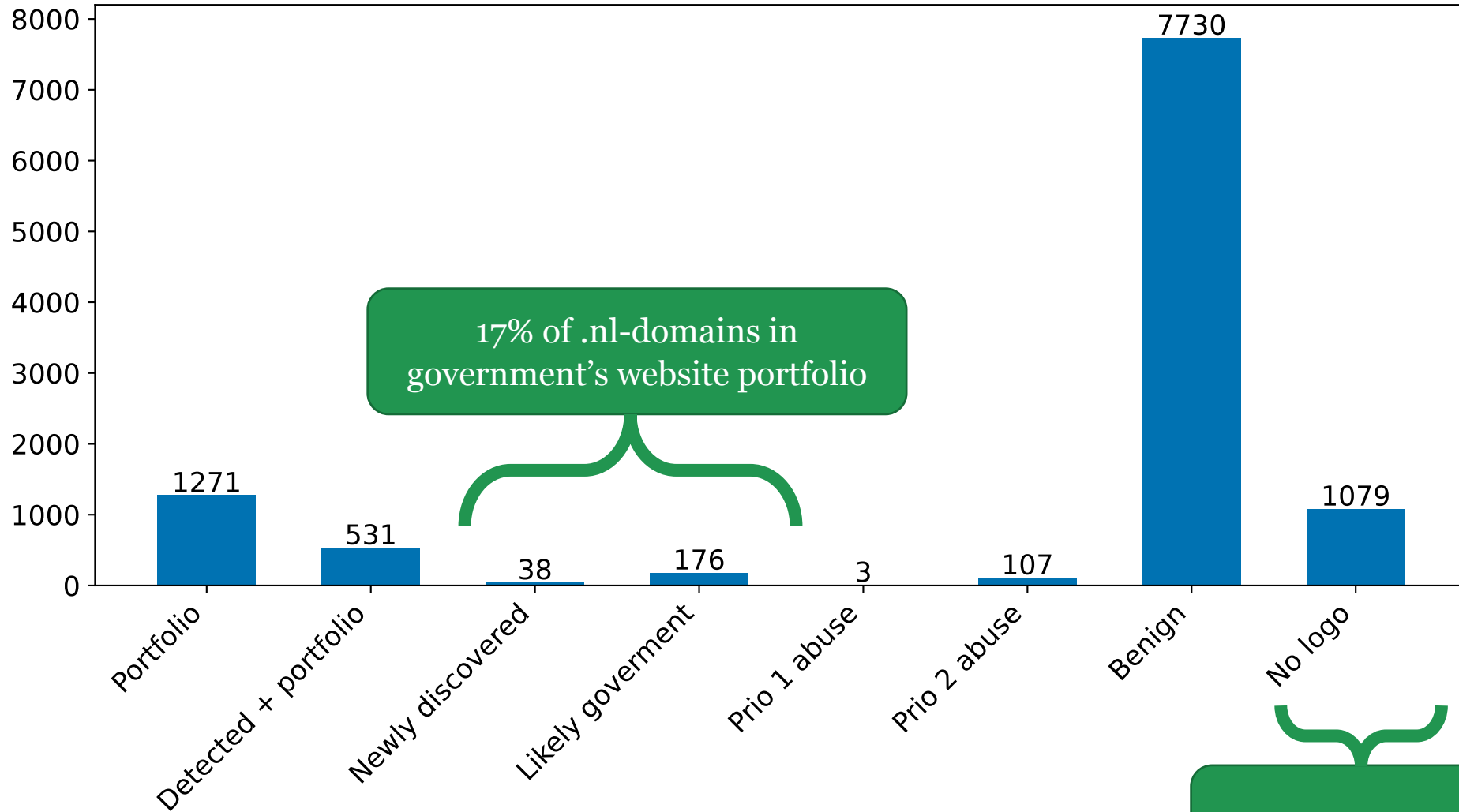
Save and exit

# Can logo detection contribute to a safe .nl-zone?

- Evaluation study with Dutch government
- Apply to .nl-zone (6.2M): discover government domains
  - Unknown domains can be transferred or cancelled by accident
  - Unknown domains cannot be monitored
- Apply to new registrations: find phishing attacks
  - Runs for two months, cannot report results yet



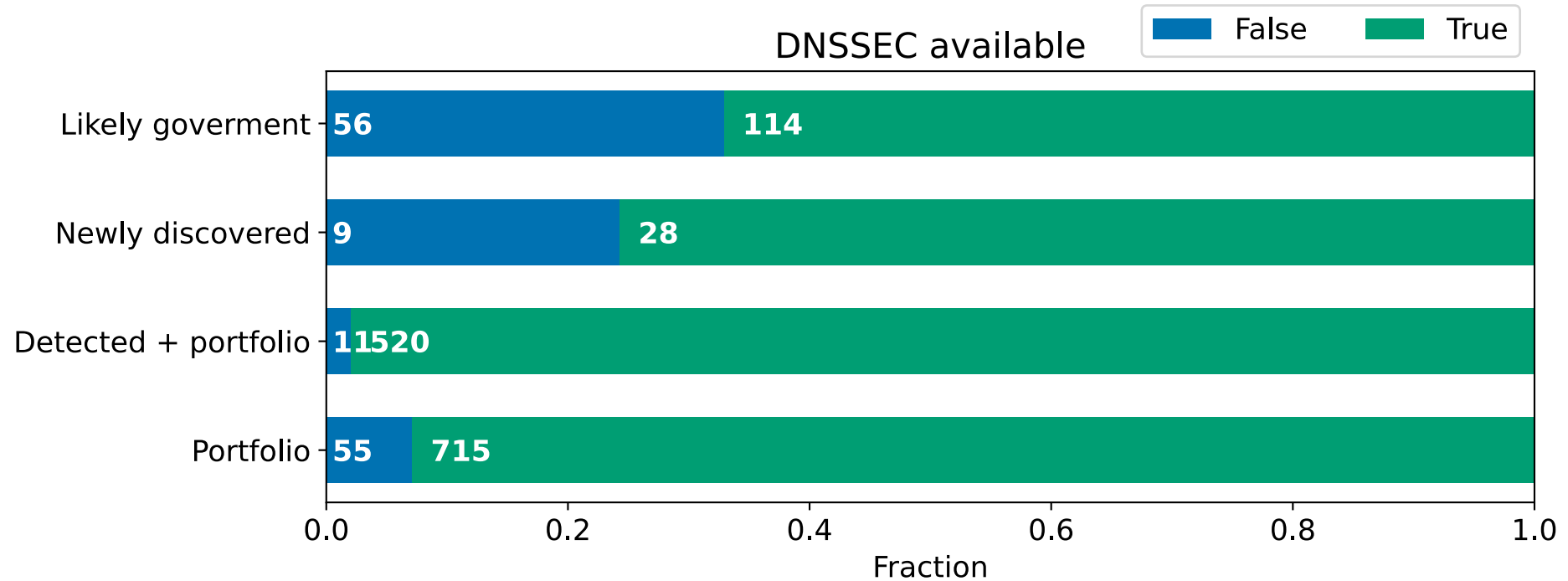
Number of domains per label (total = 10935)



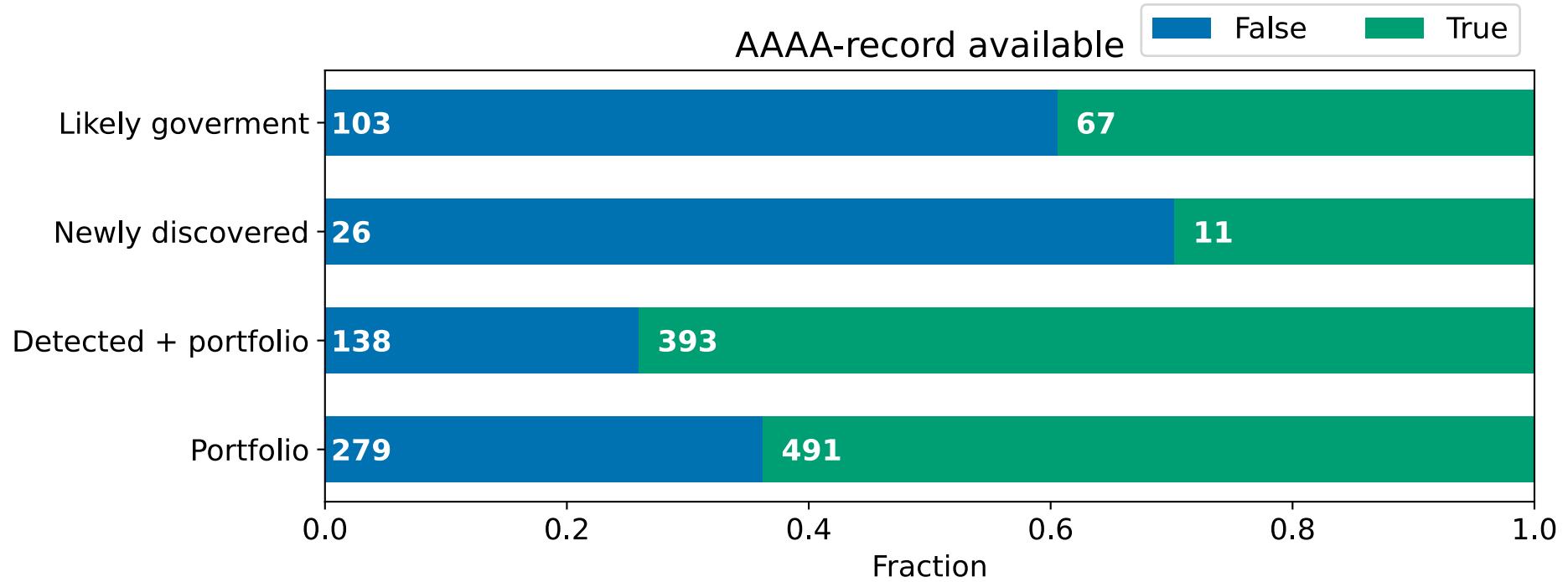
17% of .nl-domains in government's website portfolio

Precision = .89

# DNSSEC adoption per label



# IPv6 adoption per label



# Lessons learned and future work

- Visual aspects like logo's help us to detect abuse
- Logo's also help to keep domain portfolio accurate
- Large gray area of unwanted, but not abusive content

## **Next steps:**

- Pilot study with accreditation scheme for webshops
- Monitor and analyze new registrations
- Write academic paper and publish code



*Volg ons*

 SIDN.nl

 @SIDN

 SIDN

Q&A

[www.sidnlabs.nl](http://www.sidnlabs.nl) | [stats.sidnlabs.nl](http://stats.sidnlabs.nl)

Thymen Wabeke  
Research engineer  
[thymen.wabeke@sidn.nl](mailto:thymen.wabeke@sidn.nl)

