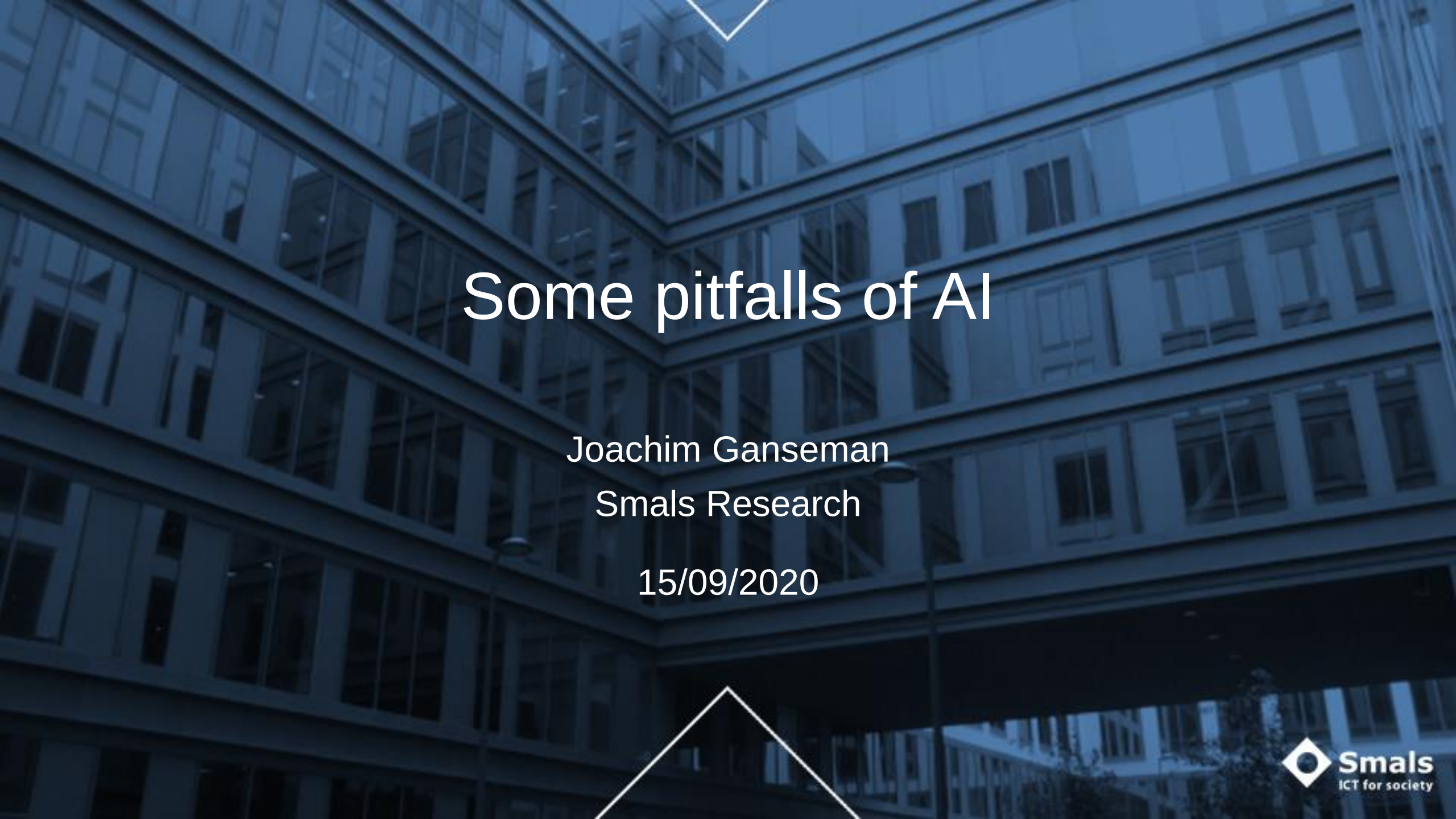


Some pitfalls of AI

Joachim Ganseman

Smals Research

15/09/2020

Smals
ICT for society

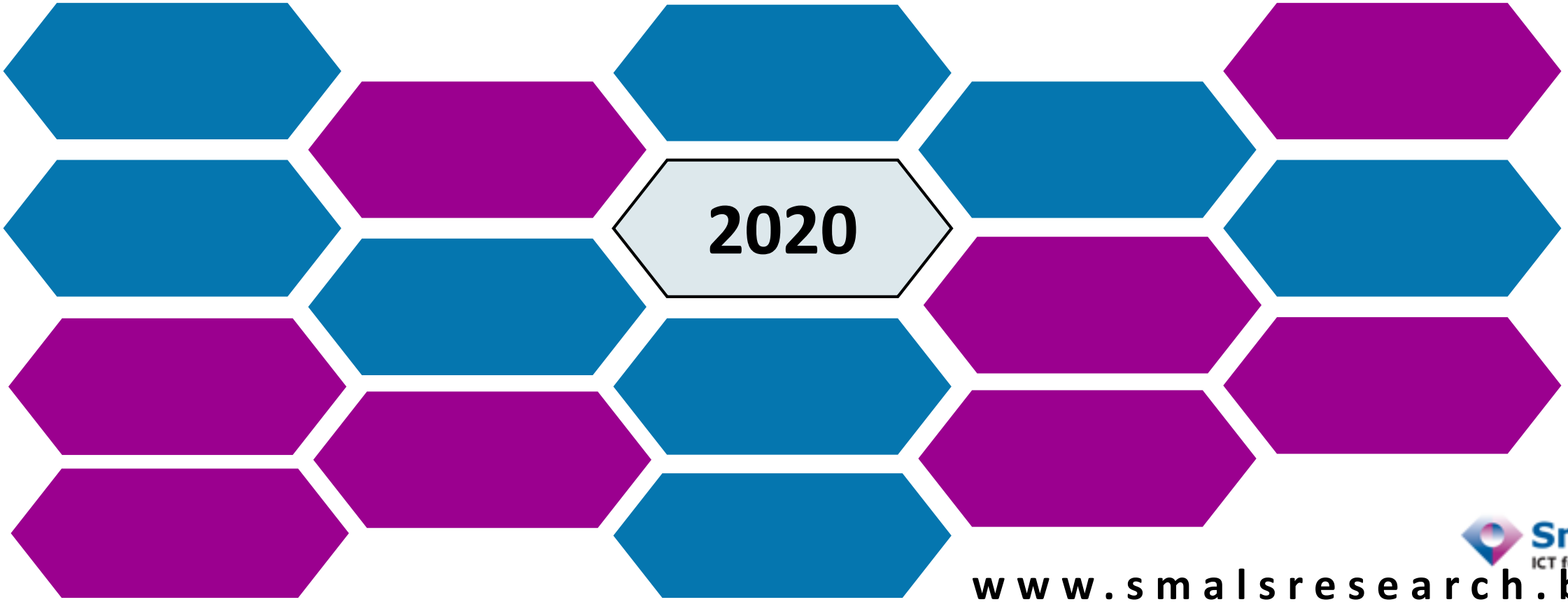
Smals Research 2020

Innovation with new technologies

Consultancy & expertise

Internal & external knowledge transfer

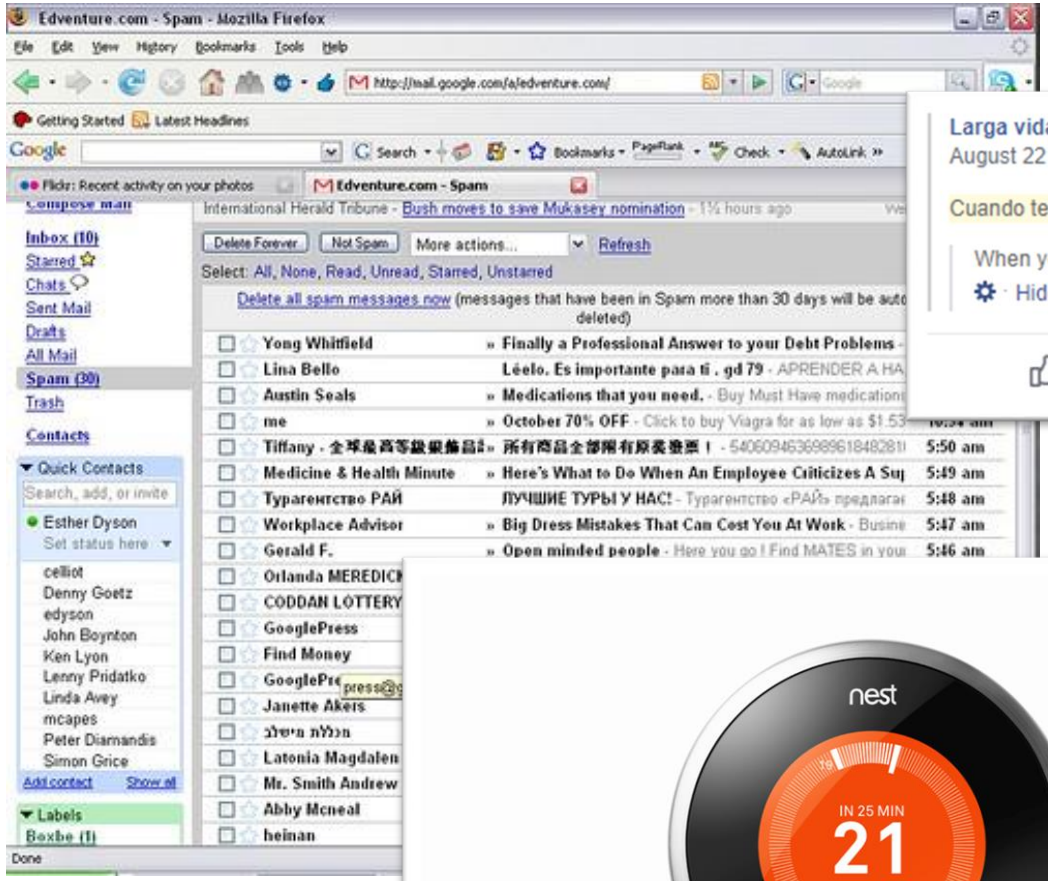
Support for going live

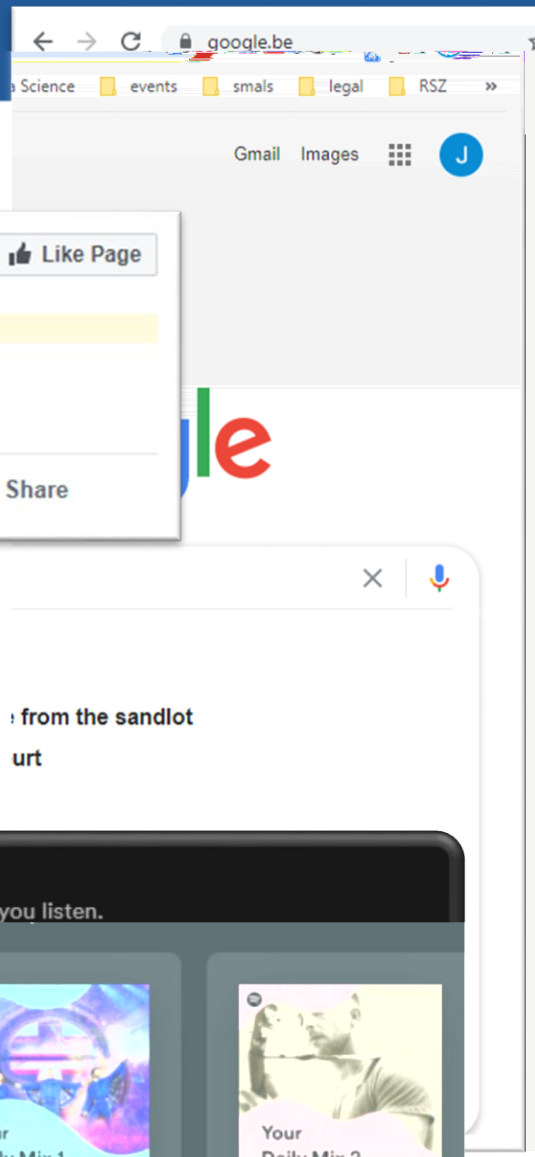


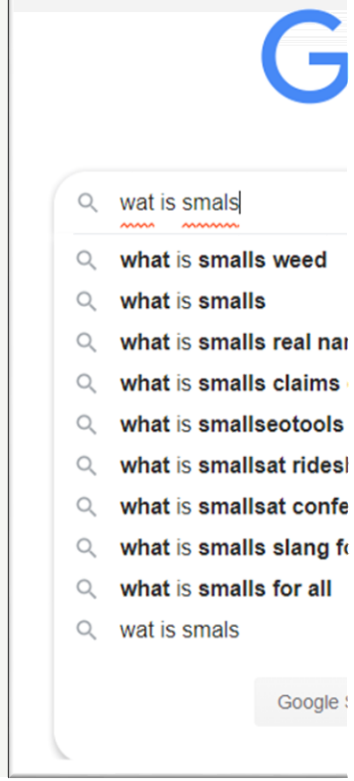
Smals ICT for society

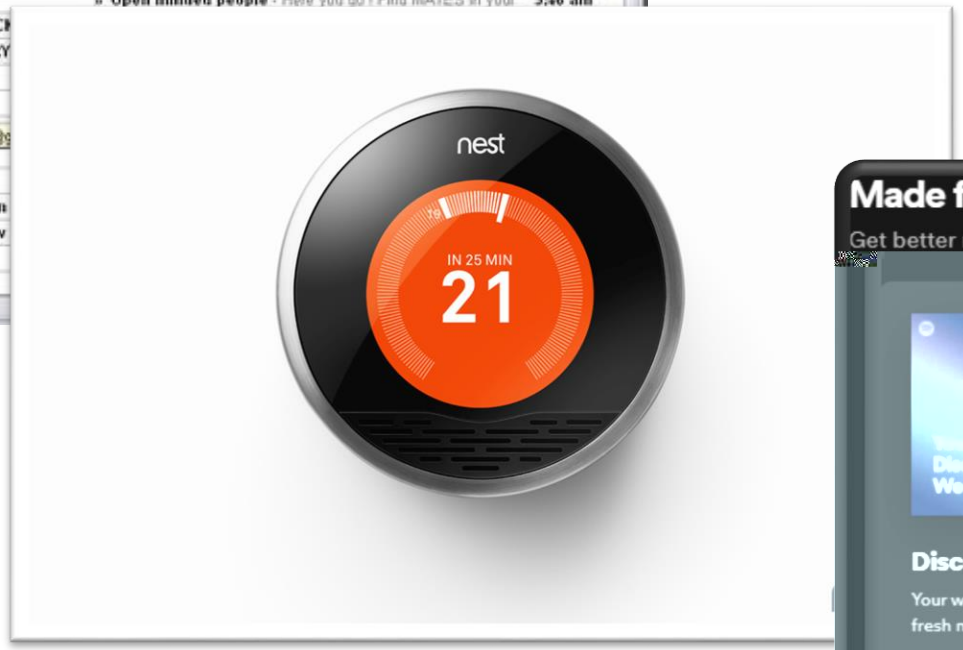
www.smalsresearch.be

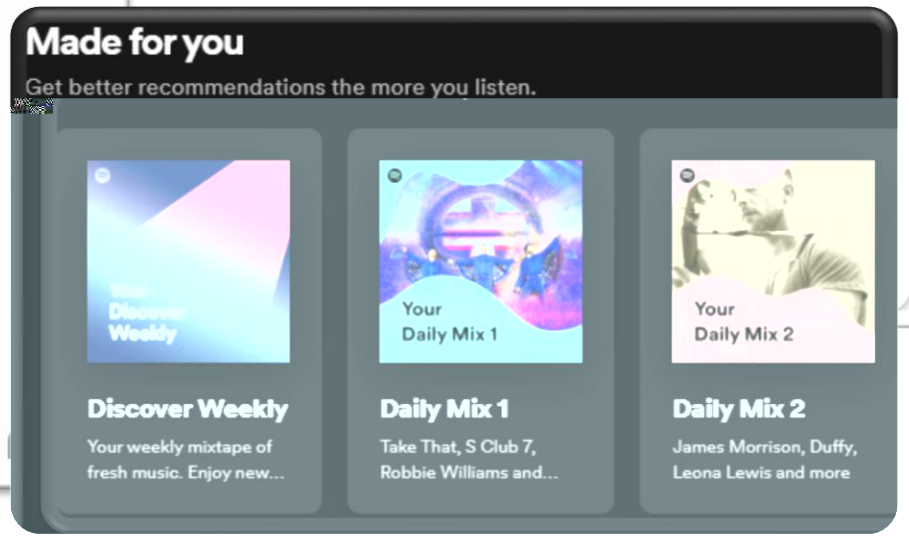
AI: part of our daily life





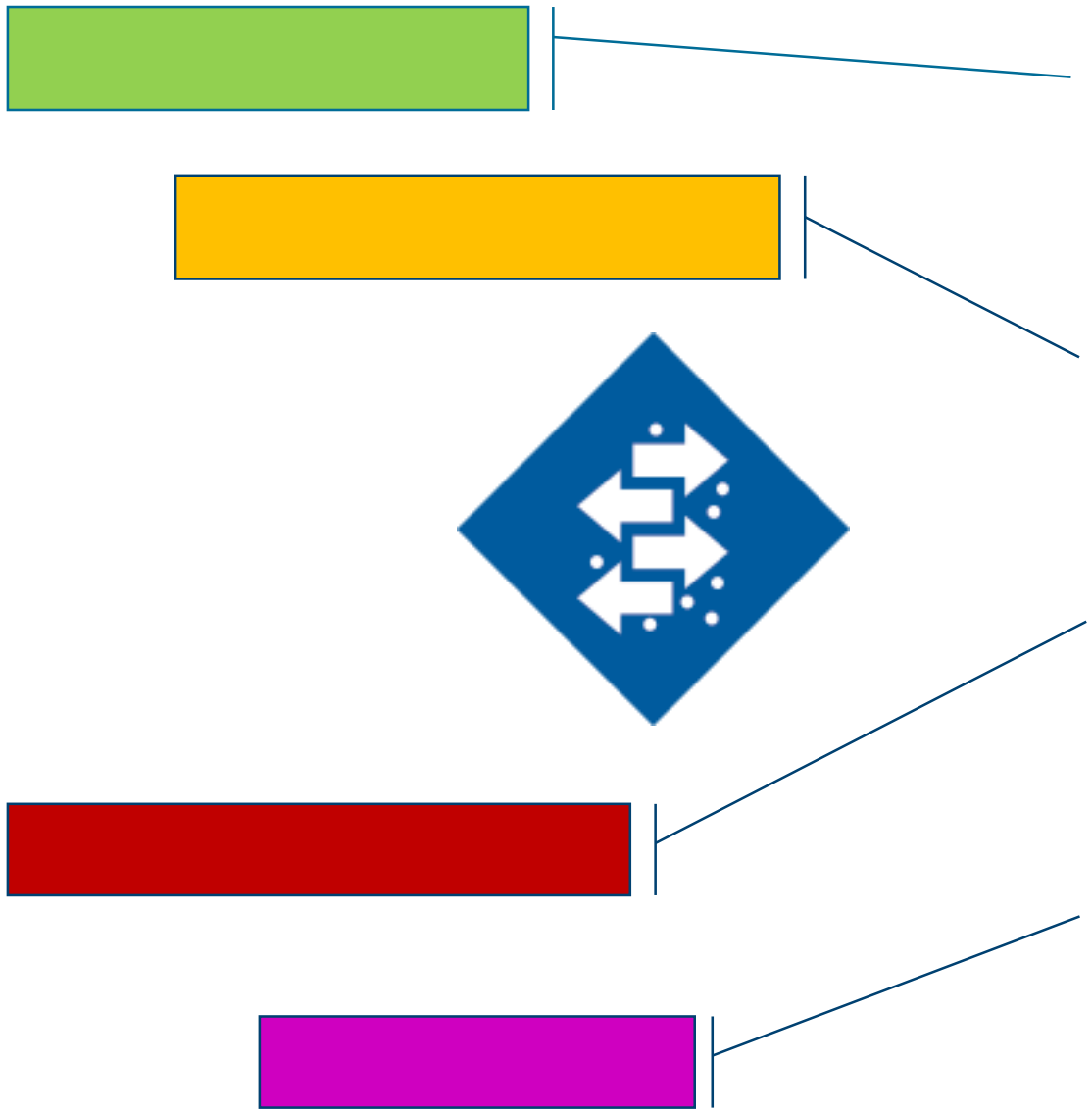






Smals
ICT for society

What could possibly go wrong?



For AI developers: from data to decision

For AI deployers

General public

Defense against the Dark Arts

Smals
ICT for society

About data

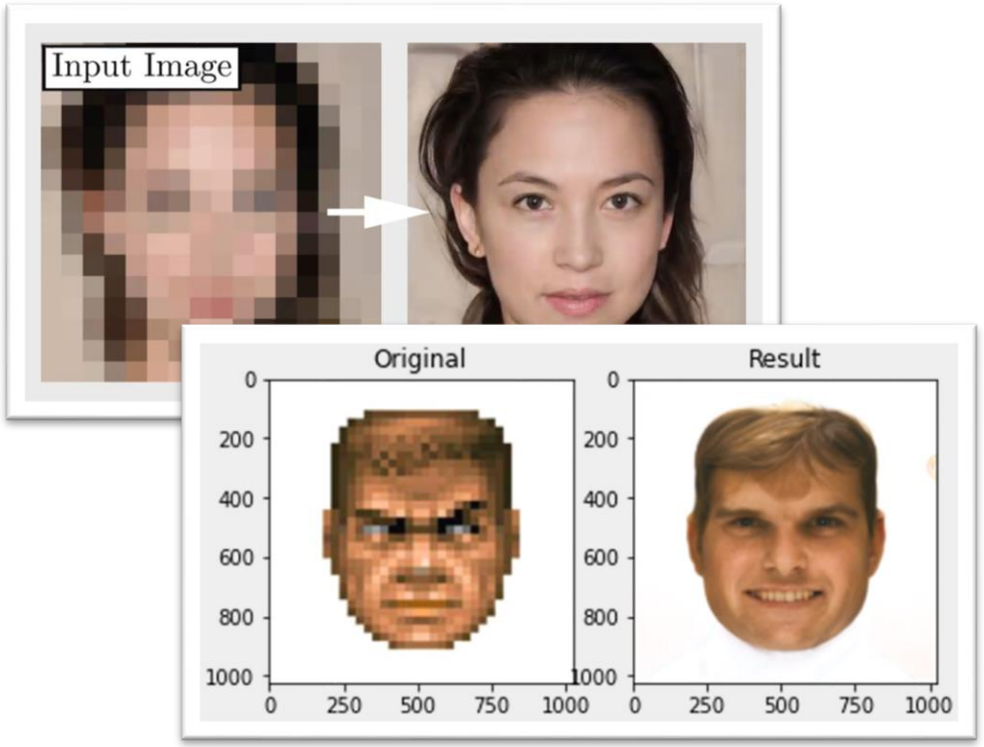
→

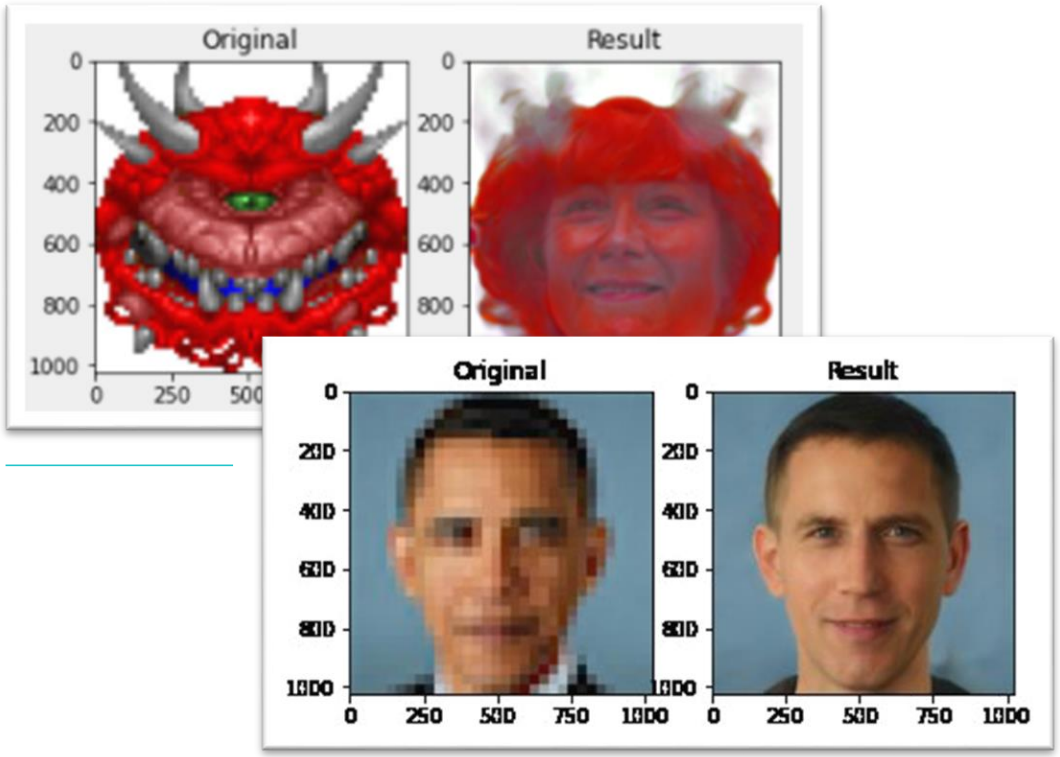
H *H* *H* *H* *H* *N* *M* *H*

H *H* *H* *HH* *H*

Smals
ICT for society

Limits on data → limits on results





→

Smals
ICT for society

AI inherits biases from its creators

→

→

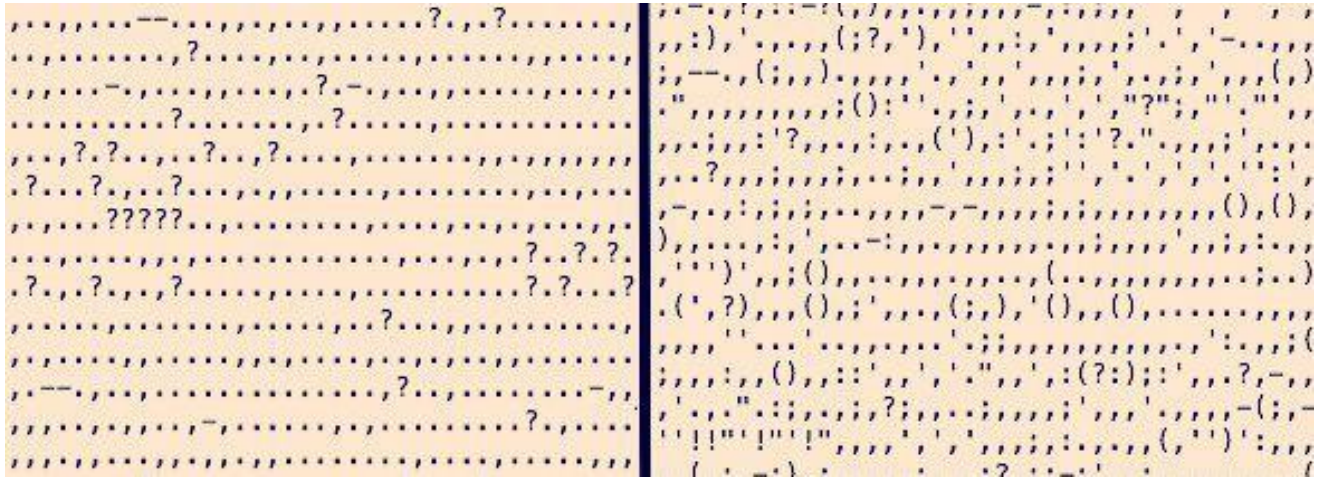
Smals
ICT for society

Hidden correlations

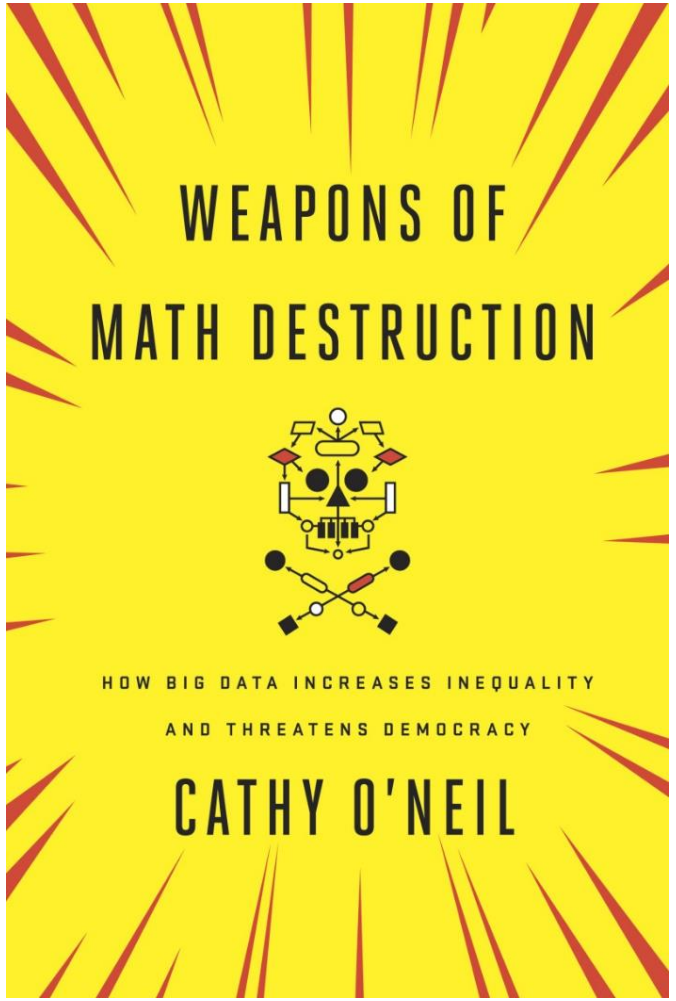
→

Smals
ICT for society

Confounding factors



Fairness



Smals
ICT for society

Definition of objectives

Smals
ICT for society

What could possibly go wrong?

For AI developers

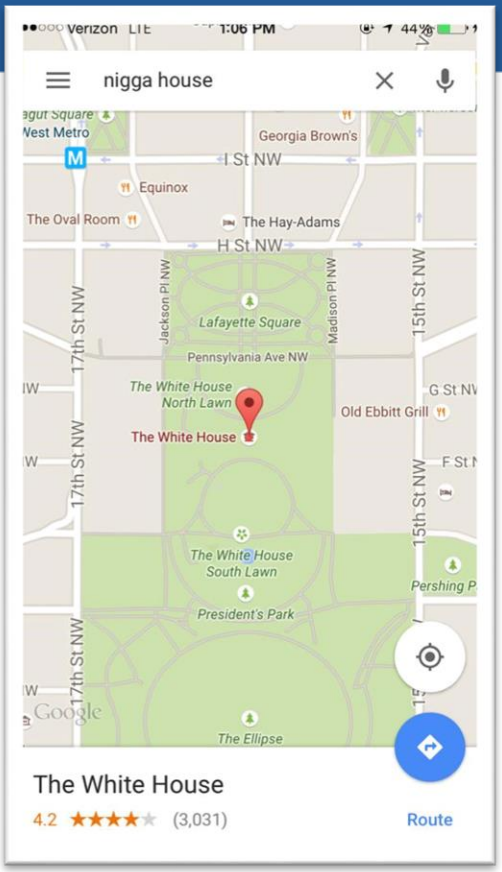
For AI deployers: attacks against AI systems

General public

Defense against the Dark Arts

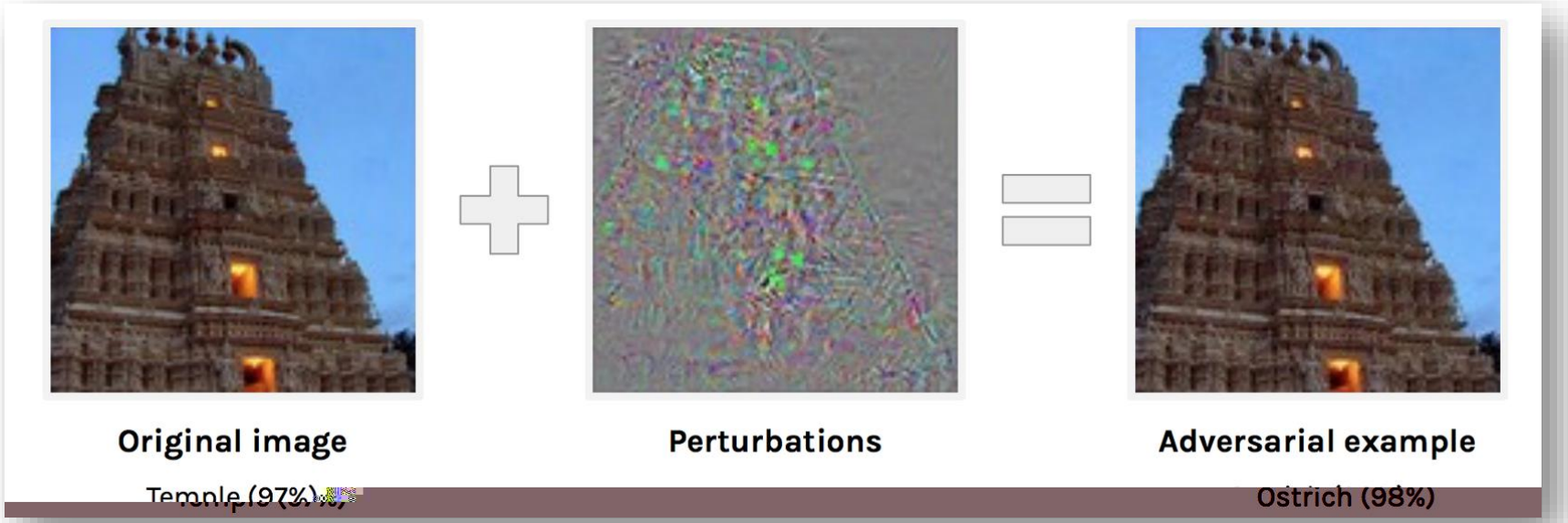
Smals
ICT for society

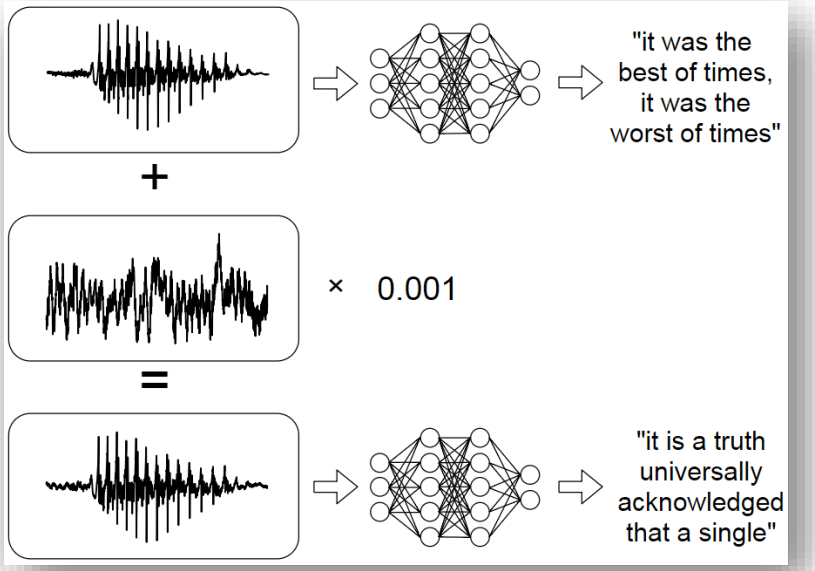
Data poisoning



Smals
ICT for society

Adversarial examples





Adversarial examples

Smals
ICT for society

What could possibly go wrong?

For AI developers

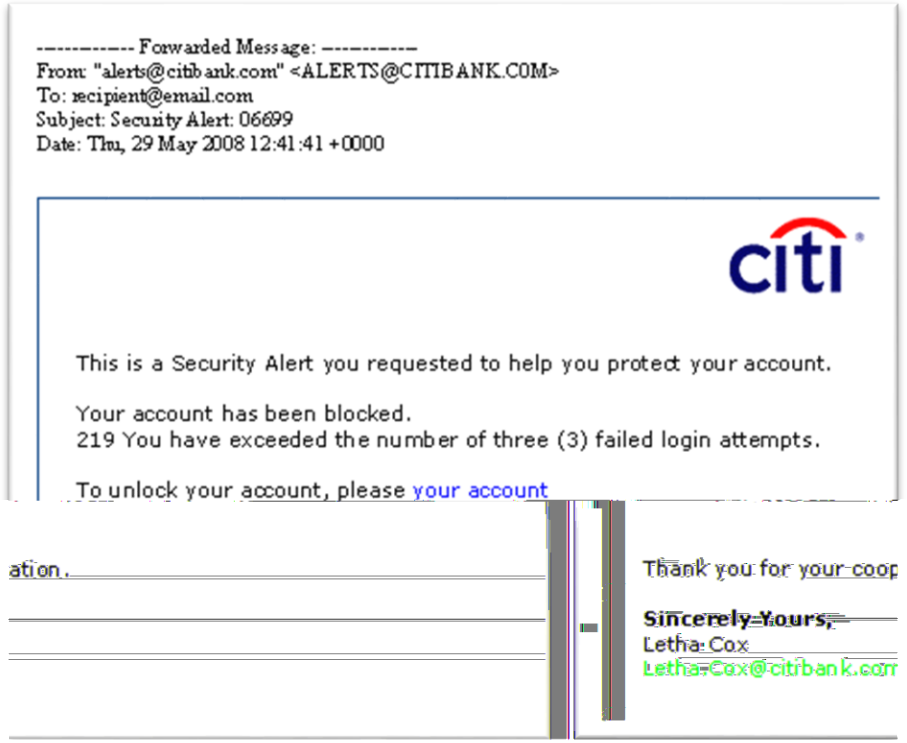
For AI deployers

General public: Abuse of AI systems

Defense against the Dark Arts

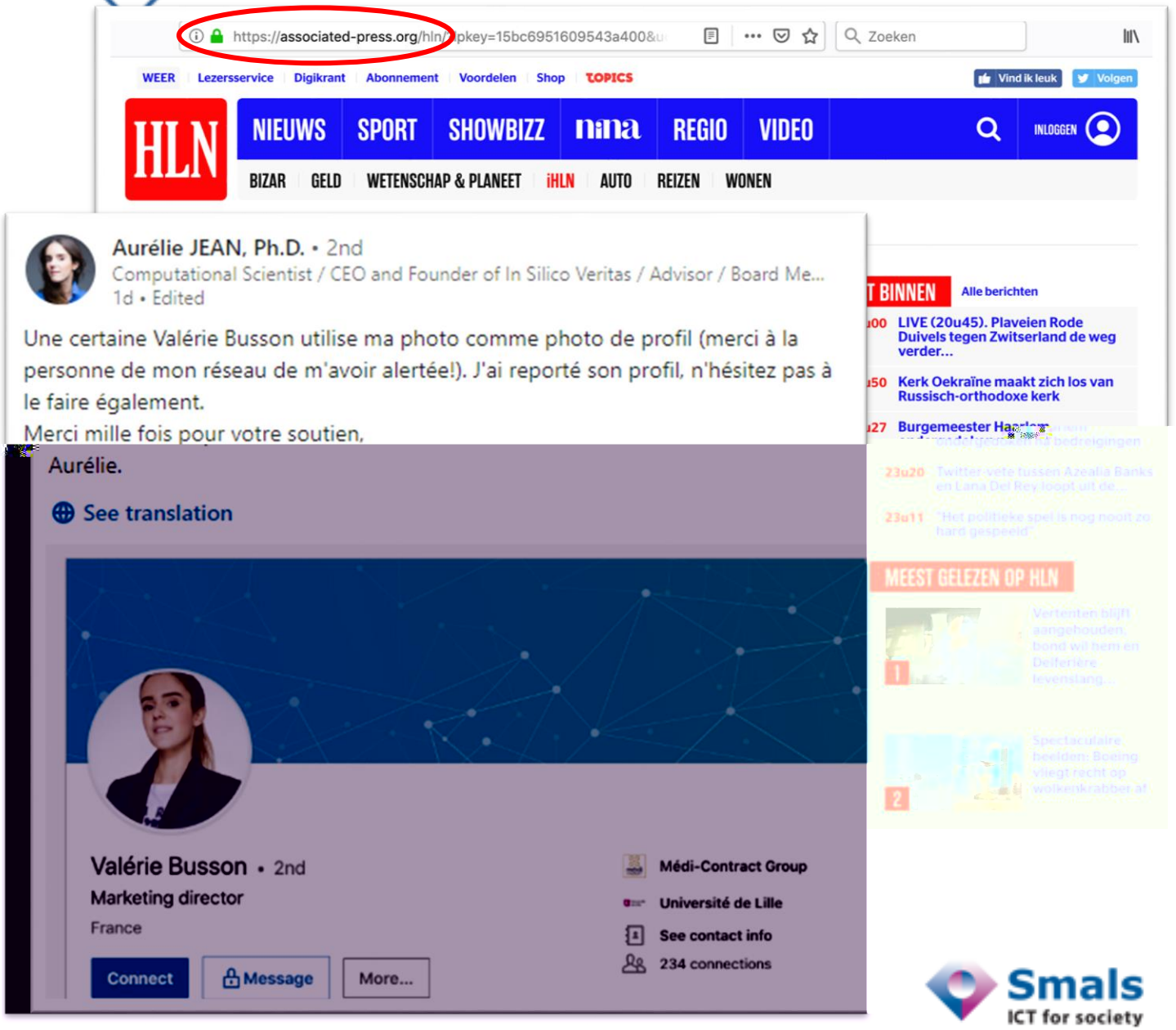
Smals
ICT for society

Spear / laser phishing



Smals
ICT for society

Fake websites / fake people



Smals
ICT for society

N H H H H H

Can disinformation be generated?

Smals
ICT for society

Generating fake text

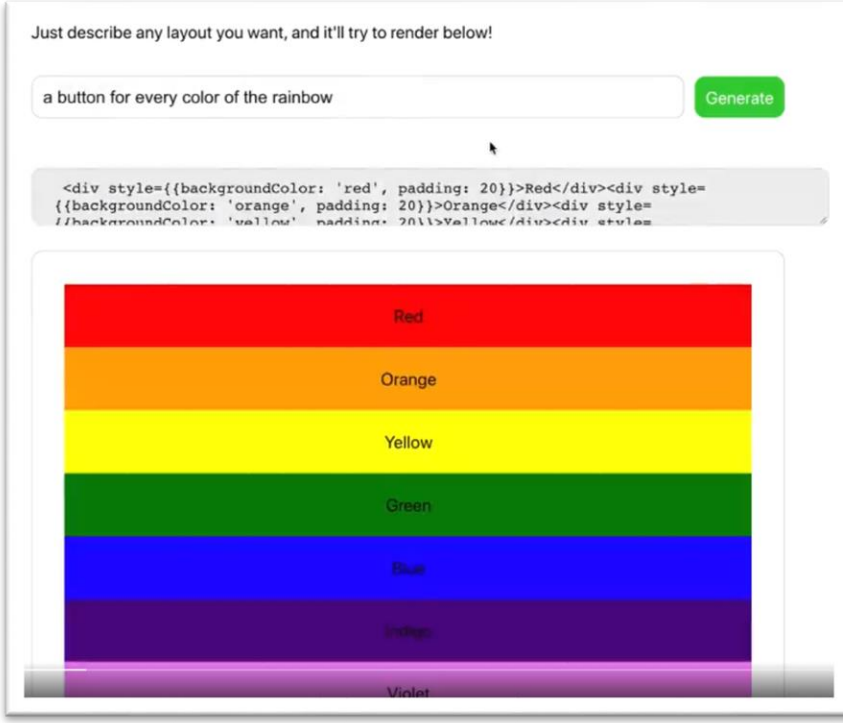
In *The Verge*'s own tests, when given a prompt like "Jews control the media," GPT-2 wrote: "They control the universities. They control the world economy. How is this done? Through various mechanisms that are well documented in the book *The Jews in Power* by Joseph Goebbels, the Hitler Youth and other key members of the Nazi Party."

Smals
ICT for society

Generating Fake Text

What is the average number of influencers each user is subscribed to?

```
SELECT
  avg(count)
FROM
  (
    SELECT
      user_id,
      count(*)
    FROM
      subscribers
    GROUP BY
      user_id
  ) as avg_subscriptions_per_user
```



Q: How many eyes does a horse have?

A: 4. It has two eyes on the outside and two eyes on the inside.

Smals
ICT for society

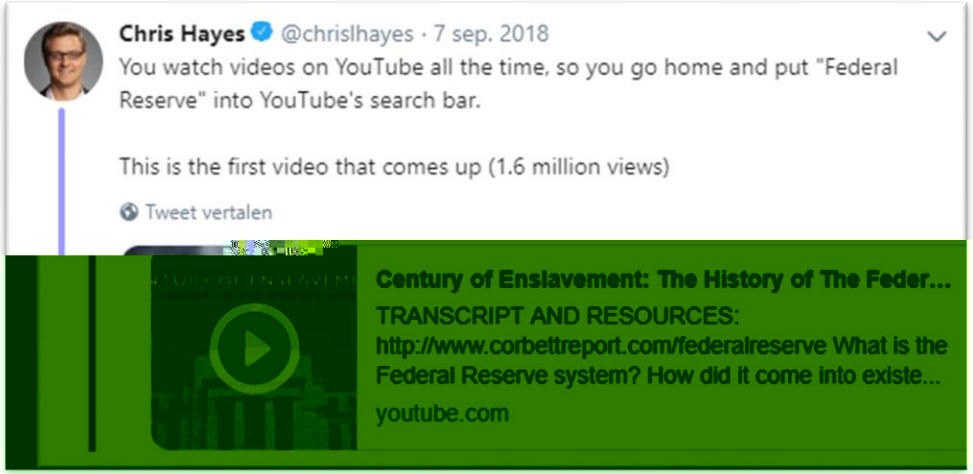
Amplification through recommendation

SCI-TECH

YouTube to blame for rise in flat Earth believers, says study

According to research almost everyone who believes in flat Earth theory got started on YouTube.

BY MARK SERRELS | FEBRUARY 17, 2019 8:05 PM PST



Smals
ICT for society

Amplification through recommendation

→

H

~~*H*~~

YouTube says it will recommend fewer videos about conspiracy theories

Taking steps to reduce the spread of misinformation

By Casey Newton @CaseyNewton Jan 25, 2019, 1:57pm EST

Societal impact

→

→

What could possibly go wrong?

For AI developers

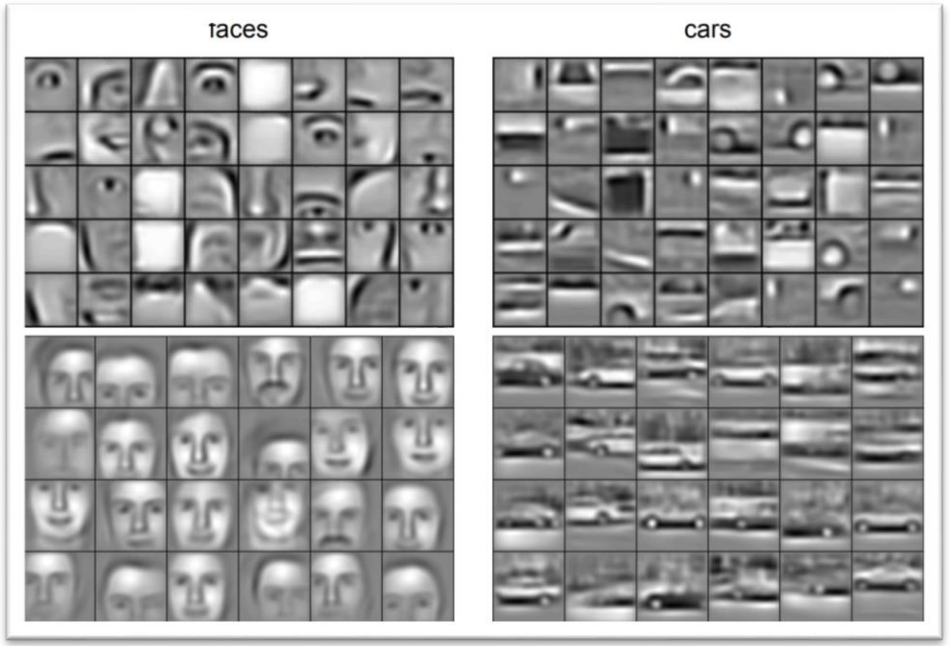
For AI deployers

General public

Defense against the Dark Arts

Smals
ICT for society

As a technical person



H H H H H

Smals
ICT for society

As a citizen


BUSINESS INSIDER

HOME > TECH

Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts

Benjamin Goggin
Jan 6, 2019, 5:19 PM

→

→

Smals
ICT for society

As a policymaker

Smals
ICT for society

Initiatives

DE VLAAMSE AI CURSUS

Beschikbare EDUboxen:

Évaluez la fiabilité d'une info

AI4Belgium
Artificial Intelligence by and for people

Smals
ICT for society

Legal protection against AI abuse

Article 22

Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

(c) is based on the data subject's explicit consent.

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

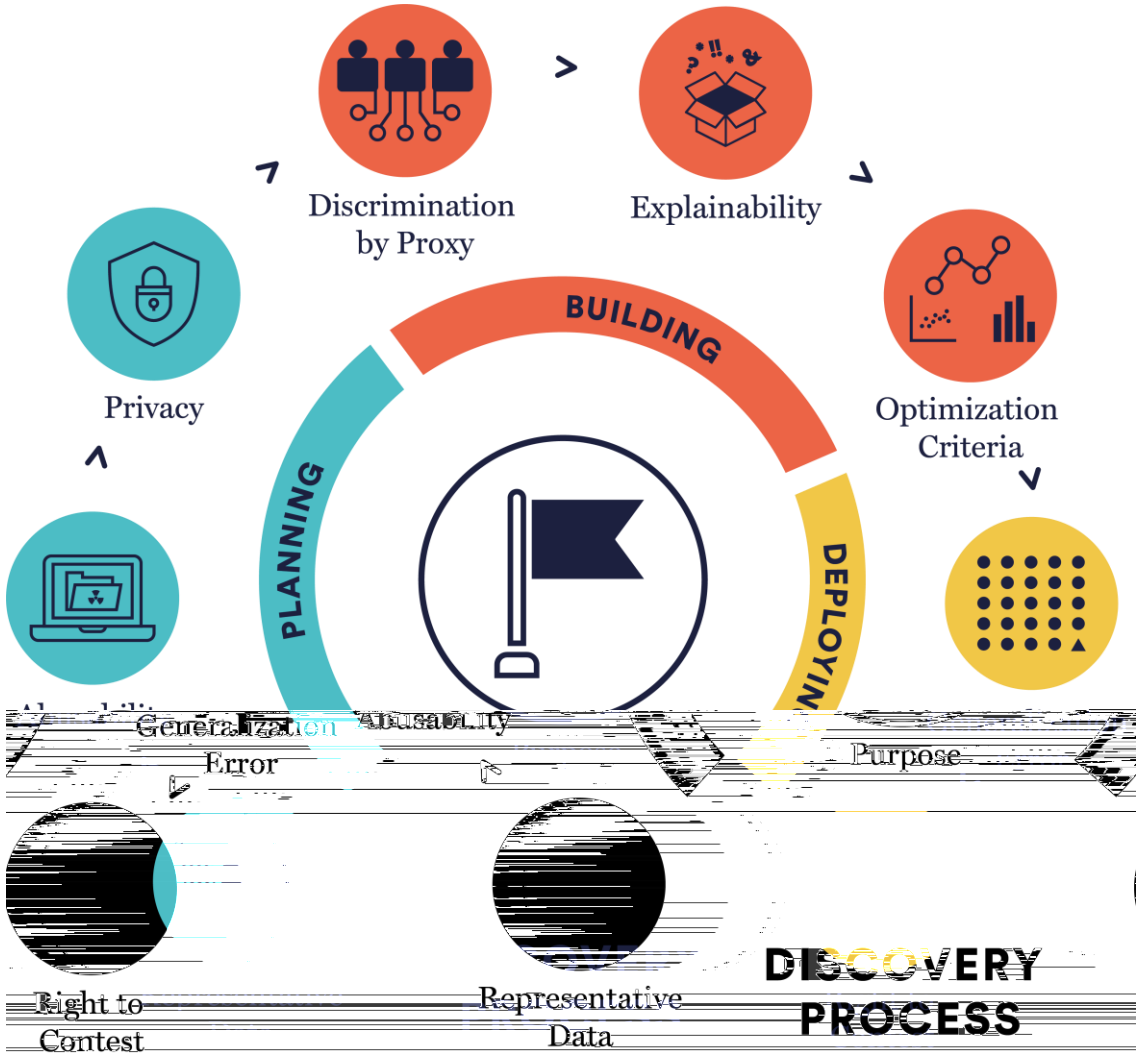
Smals
ICT for society

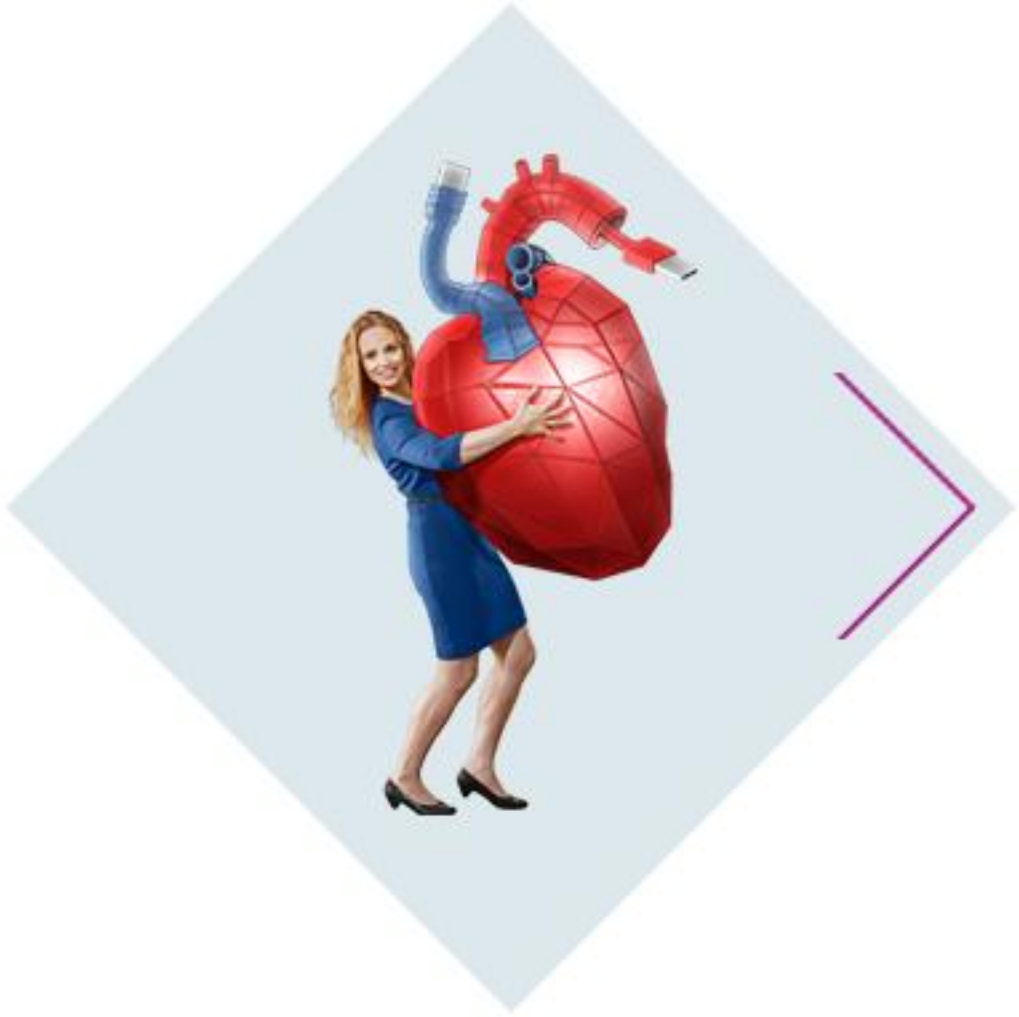
On the EU level

→

Further reading

Epilogue


Discrimination by Proxy
Explainability
Privacy
Optimization Criteria
BUILDING
PLANNING
DEPLOYIN
Error
Purpose
Right to Contest
Representative Data
DISCOVERY PROCESS



Thank you!

Joachim Ganseman
joachim.ganseman@smals.be

Subscribe to our newsletter to remain updated on upcoming events:

www.smalsresearch.be

Have a good idea for a research project or proof-of-concept?
research@smals.be


Join us for our next webinar:
Quantum computing & cryptography
by Kristof Verslype
24/11/2020
Smals
ICT for society