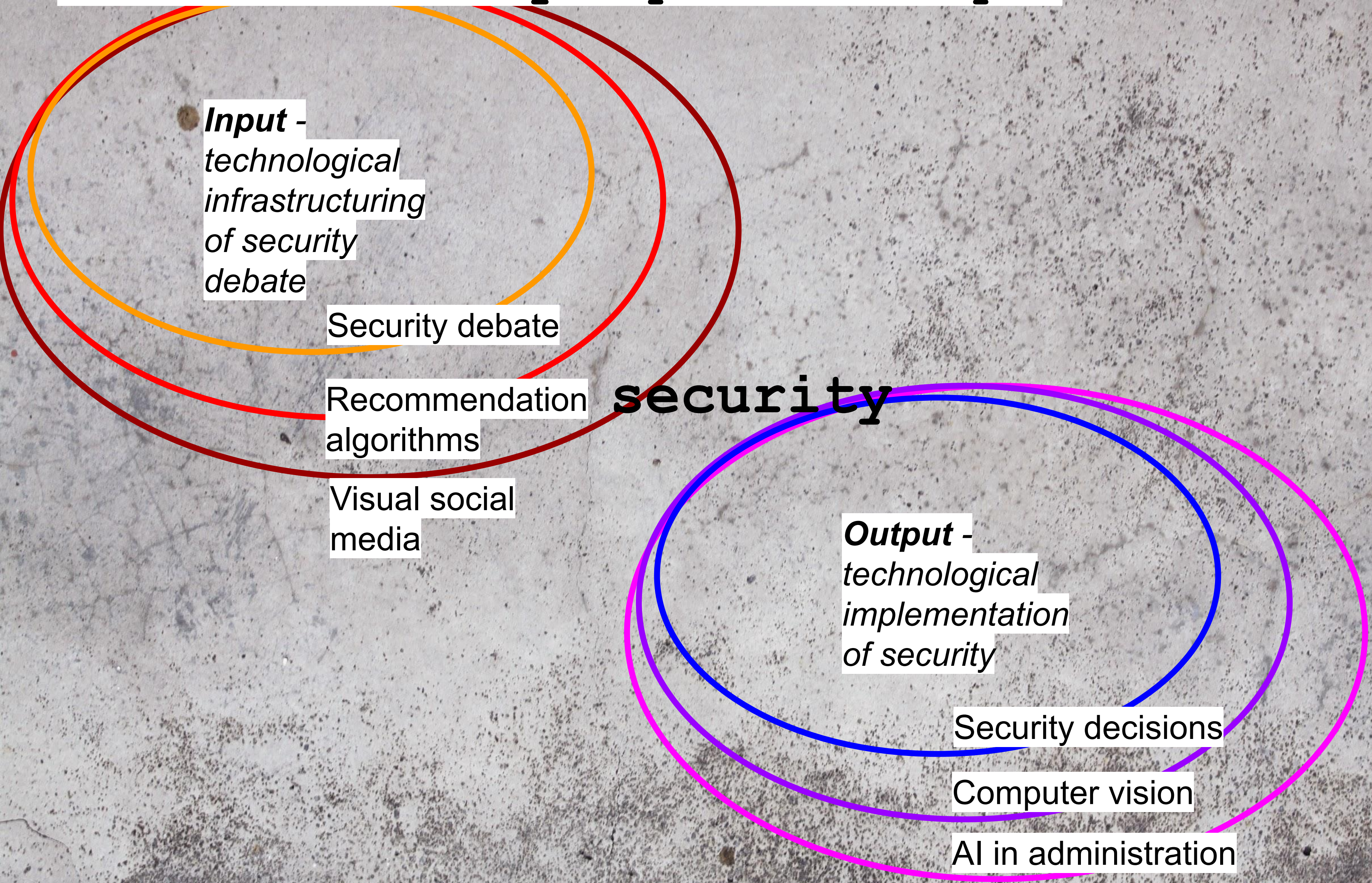
How Digitization and Social Media affects Internal Security

Rune Saugmann,
Academy of Finland Fellow
Tampere University

Saugmann.tumblr.com

How Digitization and Social Media affects Internal Security input and output



Input *- technological infrastructuring of security debate*

Security debate

Recommendation algorithms

Visual social media

security

Output *- technological implementation of security*

Security decisions

Computer vision

AI in administration

Saugmann.tumblr.com

Input to security:
Digital video in security and world politics

1991

2001

You Tube

Broadcast Yourself™

After 2007

News and fake news

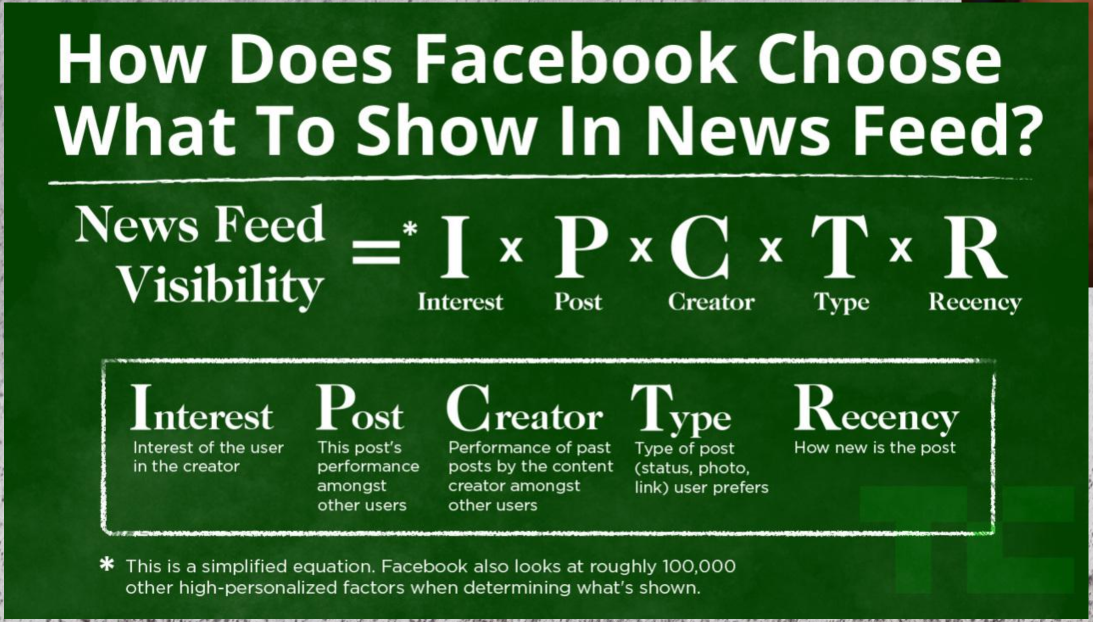
War and Propaganda

Popular mobilization

Timeline: saugmann.tumblr.com

Input to security:

- Images lead to “reality bias”,
- Social media edits and filters reality by powerful recommendation algorithms



How Does Facebook Choose What To Show In News Feed?

News Feed Visibility =* I x P x C x T x R

Interest · Post · Creator · Type · Recency

Interest Interest of the user in the creator

Post This post's performance amongst other users

Creator Performance of past posts by the content creator amongst other users

Type Type of post (status, photo, link) user prefers

Recency How new is the post

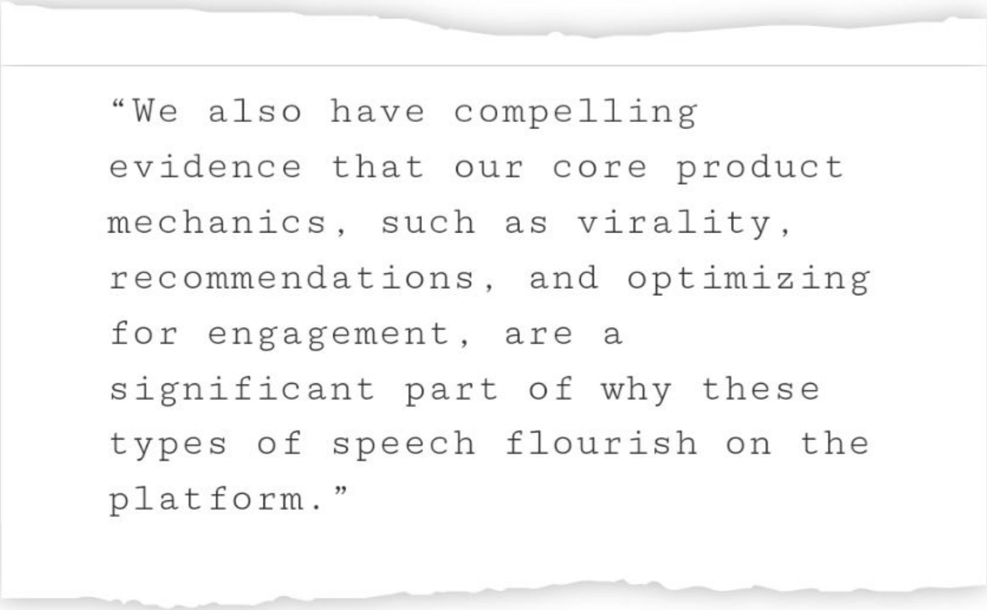
* This is a simplified equation. Facebook also looks at roughly 100,000 other high-personalized factors when determining what's shown.

Saugmann.tumblr.com

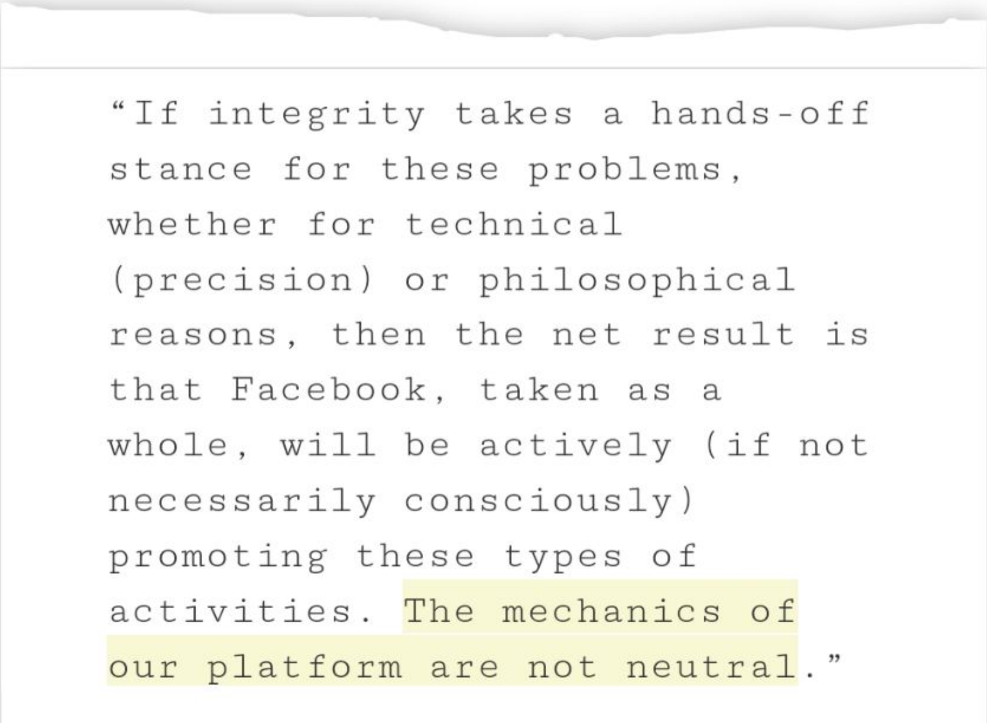
Input to security:

Facebook leaks of internal research (2021)

-> conflict and conspiracy bias in the media citizens use to talk to each other



"We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform."



"If integrity takes a hands-off stance for these problems, whether for technical (precision) or philosophical reasons, then the net result is that Facebook, taken as a whole, will be actively (if not necessarily consciously) promoting these types of activities. The mechanics of our platform are not neutral."

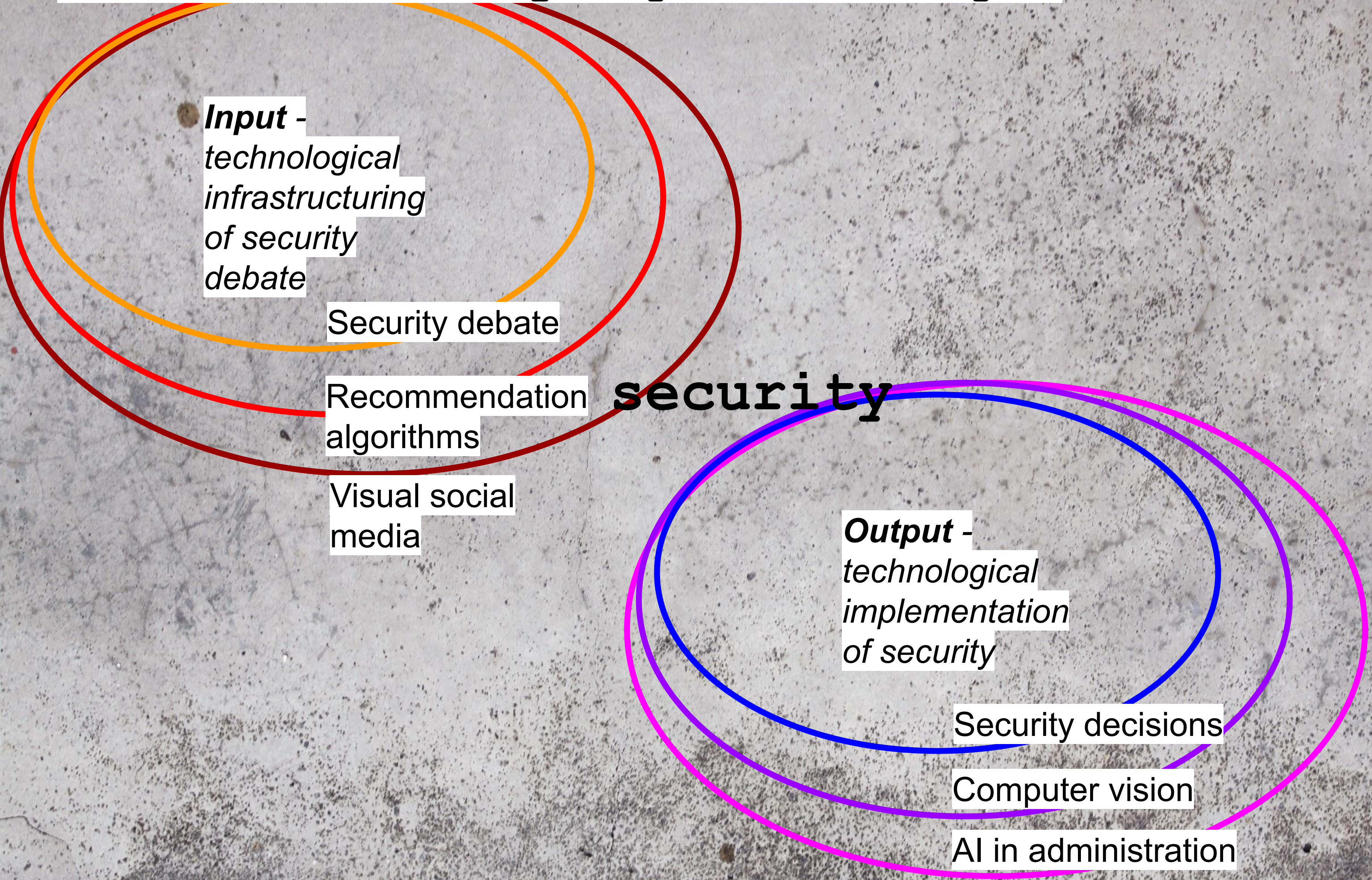
Input to security recap: QUESTIONS TO THINK ABOUT

- ‘Reality bias’- images seem to show reality unmediated
- Conflict bias - social media promotes conflict

- What can you do to work proactively with reality and conflict biases in our visual / social media?
  a. If you are a frontline worker, working directly with citizens

  b. What can we do as a society?

How Digitization and Social Media affects Internal Security input and output



Input *- technological infrastructuring of security debate*

Security debate

Recommendation algorithms

Visual social media

security

Output *- technological implementation of security*

Security decisions

Computer vision

AI in administration

Saugmann.tumblr.com

Output from security politics:

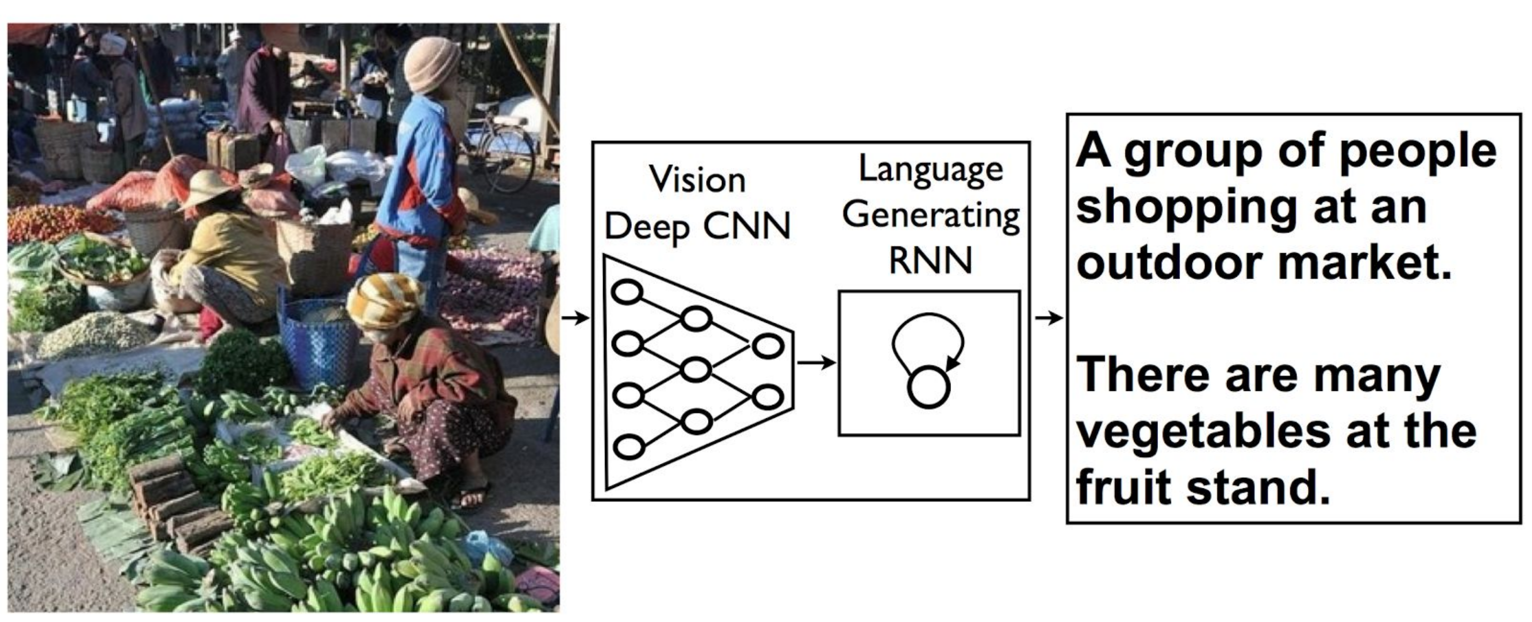
AI, computer vision and everyday security decision-making

Automation bias - trusting machines more than warranted
(Bruner & Tagiuri 1954)

Saugmann.tumblr.com

Output from security politics:

AI, computer vision and everyday security decision-making



Vision Deep CNN

Language Generating RNN

A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

Training database is the foundation of computer vision

- Has labelled data (images with descriptions)
- Becomes ‘ground truth’ for AI learning
- Difficult and expensive

Saugmann.tumblr.com

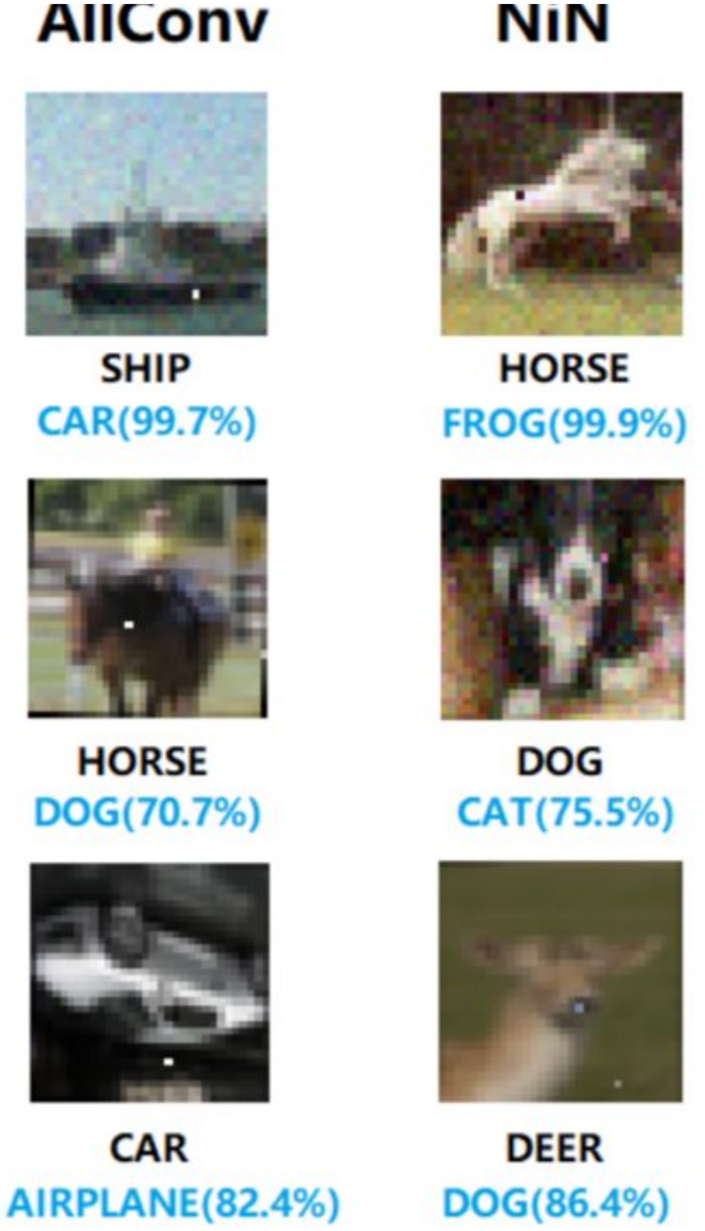
Output from security politics:

Strange results in computer vision (adversarial attacks)

Su, Vargas & Sakurai (2018): One-pixel-attacks

Adversarial attacks (Xie et al 2018):

"The success of adversarial attacks leads to security threats in real-world applications of convolutional networks, but equally importantly, it demonstrates that these networks perform **computations that are dramatically different from those in human brains.**"



Output from security politics:

Strange results in computer vision (adversarial attacks)

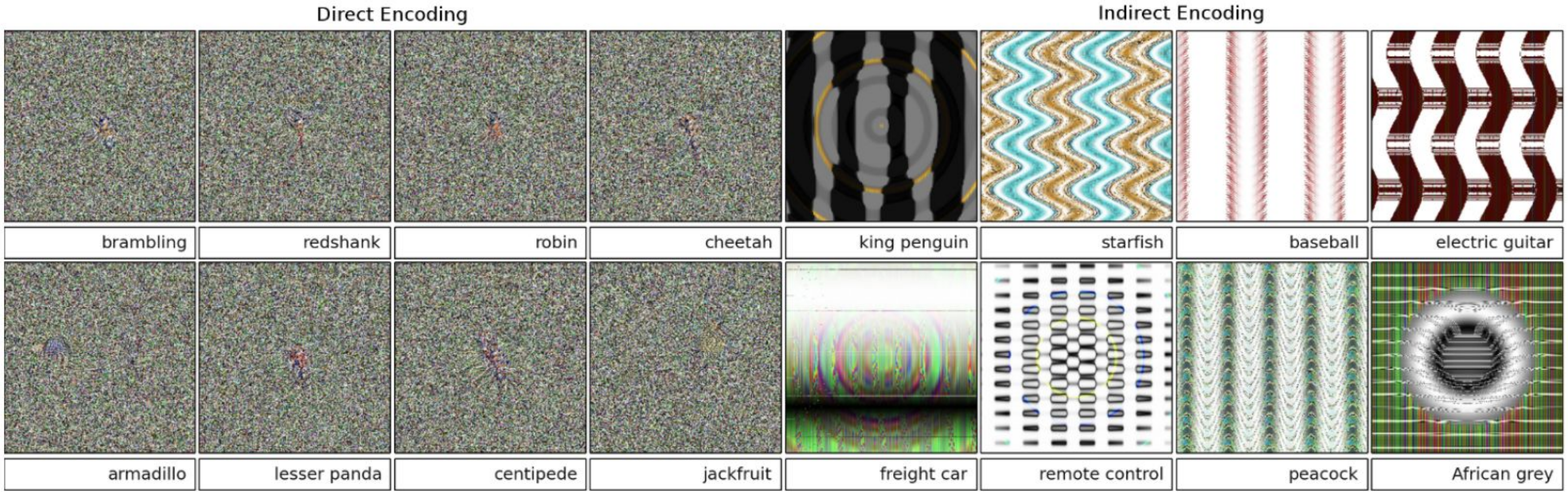
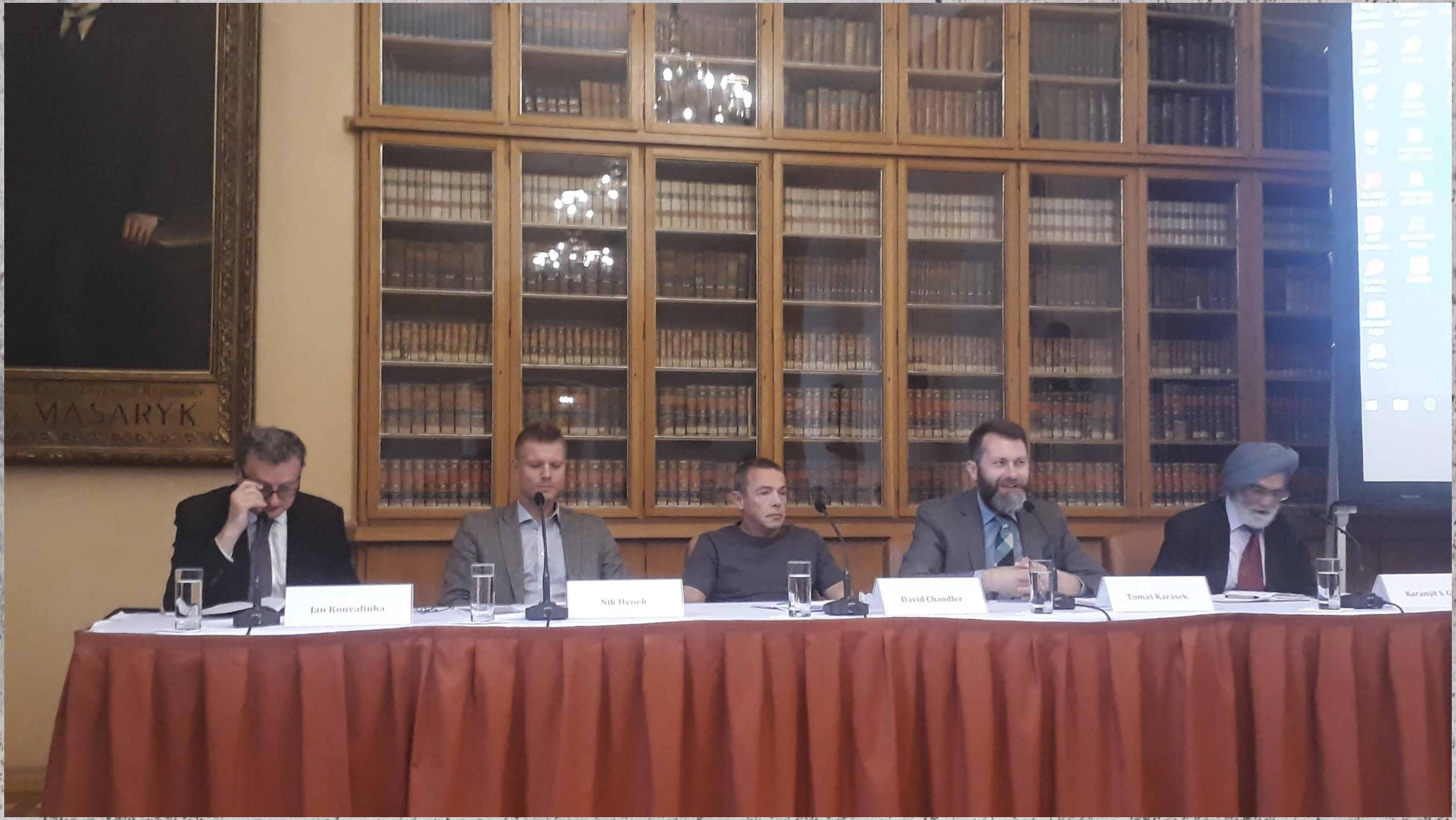


Figure 1: Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with >= 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Left: Directly encoded images. Right: Indirectly encoded images.

Saugmann.tumblr.com

Output from security politics:

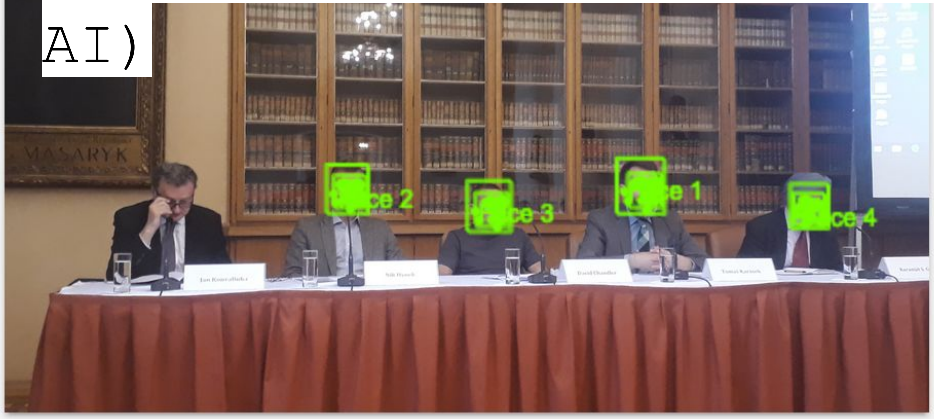
Bias in practice (experiment with Google Vision AI)



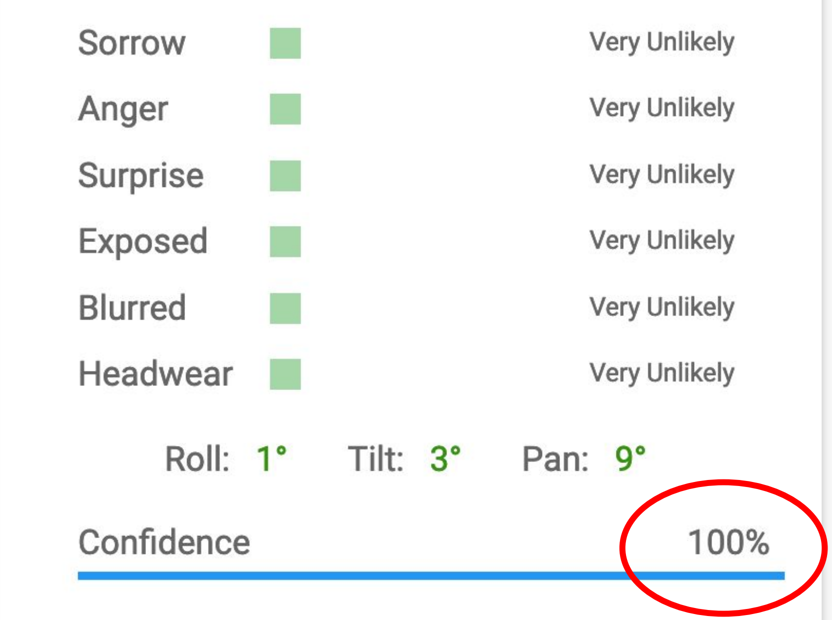
Saugmann.tumblr.com

Output from security politics:

Bias in practice (experiment with Google Vision AI)



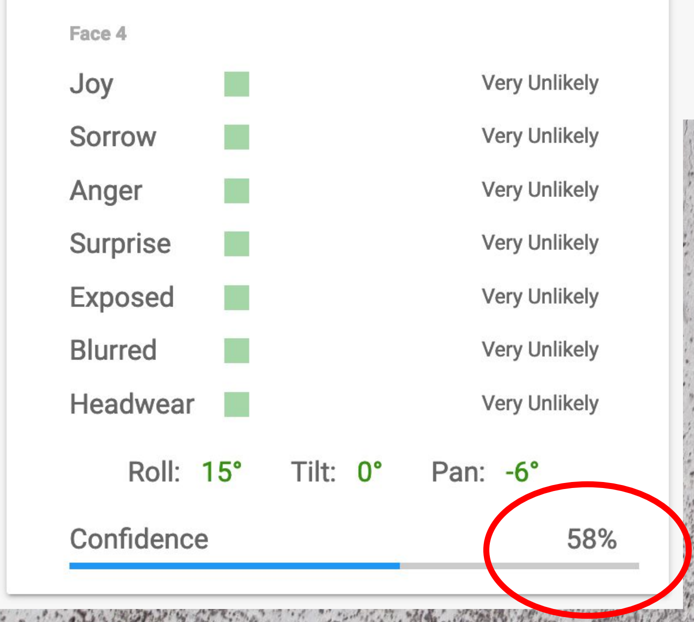
opening panel.jpg



Sorrow	🟩	Very Unlikely
Anger	🟩	Very Unlikely
Surprise	🟩	Very Unlikely
Exposed	🟩	Very Unlikely
Blurred	🟩	Very Unlikely
Headwear	🟩	Very Unlikely

Roll: 1° Tilt: 3° Pan: 9°

Confidence 100%



Face 4

Joy	🟩	Very Unlikely
Sorrow	🟩	Very Unlikely
Anger	🟩	Very Unlikely
Surprise	🟩	Very Unlikely
Exposed	🟩	Very Unlikely
Blurred	🟩	Very Unlikely
Headwear	🟩	Very Unlikely

Roll: 15° Tilt: 0° Pan: -6°

Confidence 58%

Different degrees of confidence often reflect systematic biases shown to be racialised, gendered, culturally dependent.

Is equal treatment possible?

Saugmann.tumblr.com

Output from security recap: KEY POINTS, QUESTIONS TO THINK ABOUT

- **Automation bias -** social media promotes conflict
- **'Reality bias' continues -** images seem to show reality unmediated, also with computer vison
- **Racial and minority bias -** technologies work best on majority population (representatin in training data)

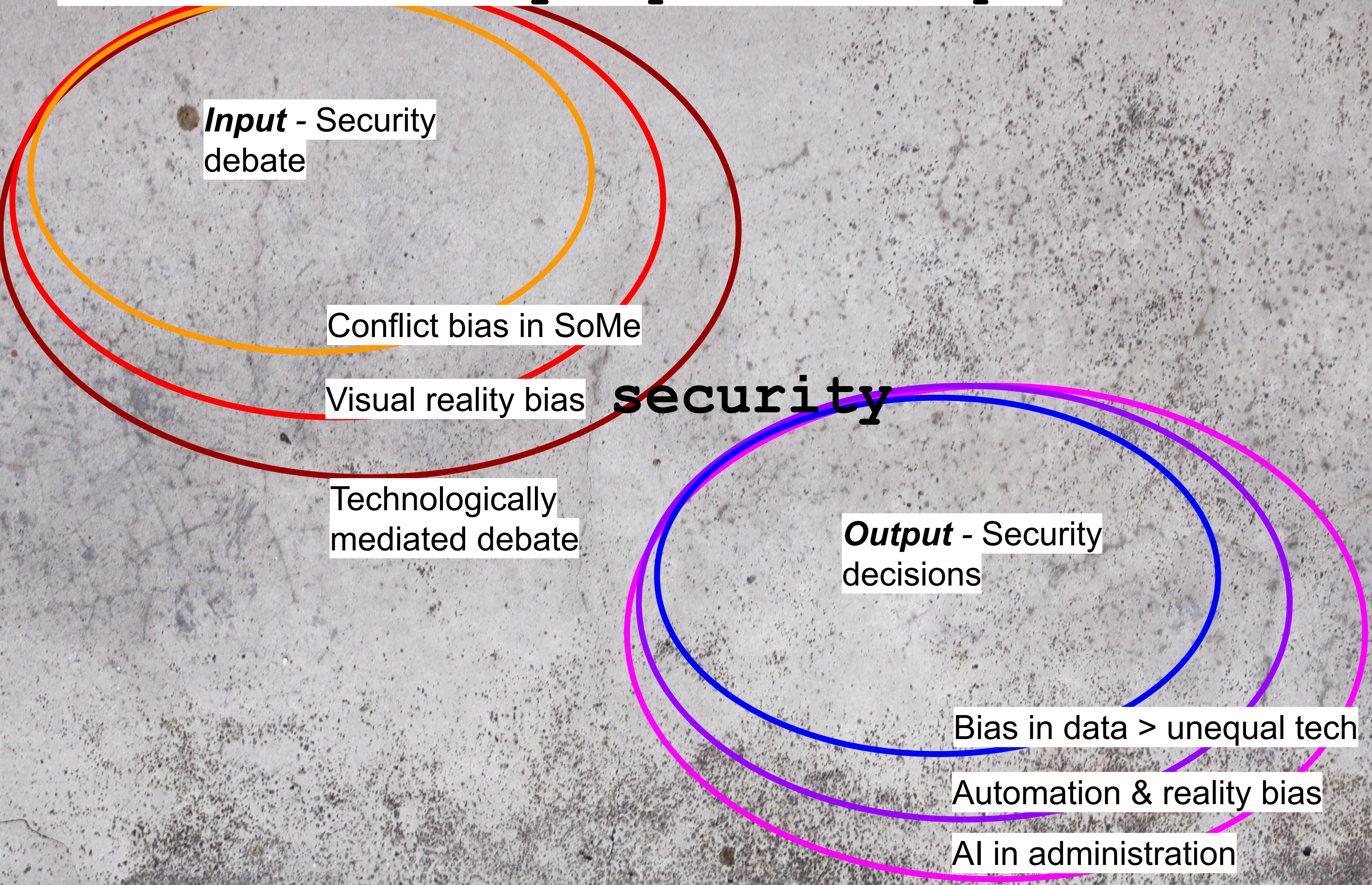
- Potential equality before the state/law problems!

Q1: If you are a frontline worker, working directly with citizens, how can you remember and counter-act biases (against minorities, in favour of tech)?

Q2: The state: How can we test, evaluate, counteract technological biases?

Saugmann.tumblr.com

How Digitization and Social Media affects Internal Security input and output



Input - Security debate

Conflict bias in SoMe

Visual reality bias

security

Technologically mediated debate

Output - Security decisions

Bias in data > unequal tech

Automation & reality bias

AI in administration

Saugmann.tumblr.com