CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

Insights on Client-Side Scanning and Alternatives in the Fight Against CSAE

Carolyn Guthoff | DeepSec 2024 | 22.11.2024 | Vienna, Austria

Carolyn Guthoff

- currently Doctoral Researcher at CISPA Helmholtz Center for Information Security
- previously Application Owner and Business Analyst at Mercedes-Benz AG
- B.Sc. and M.Sc. in Computer Science from Saarland University

Content Warning

This talk will include mention of child sexual abuse (CSA), child sexual exploitation (CSE), child sexual abuse material (CSAM) and suicide.

32,059,029

number of reports the CyberTipline of the U.S. National Center for Missing and Exploited Children (NCMEC) received in 2022

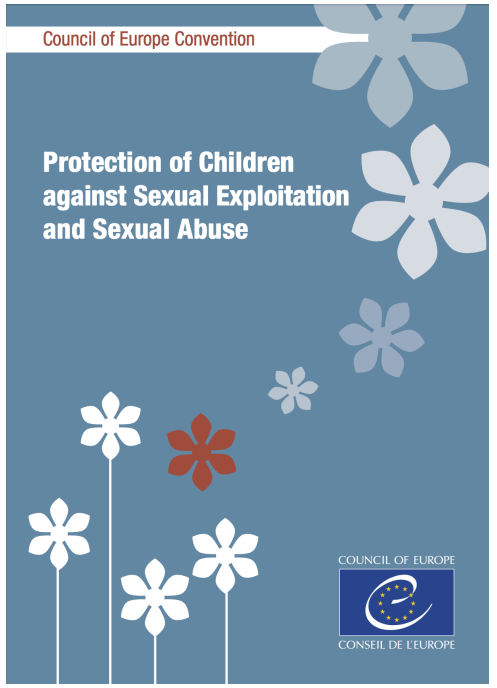
> 99.5 %

were classified as child sexual abuse material (CSAM)

CSAE and CSAM

Child Sexual Abuse and Exploitation (CSAE)

- Child Sexual Abuse (CSA) [1,2]
  - engagement in sexual activities with a child that has not reached the legal age for sexual activities
    (exclusion: sexual activities between minors)
  - engagement in sexual activities with a child
    - through coercion, force or threats
    - through abuse of a position of trust, authority or influence over the child
    - through abuse of a particularly vulnerable situation of the child, notably mental or physical disability or a situation of dependence



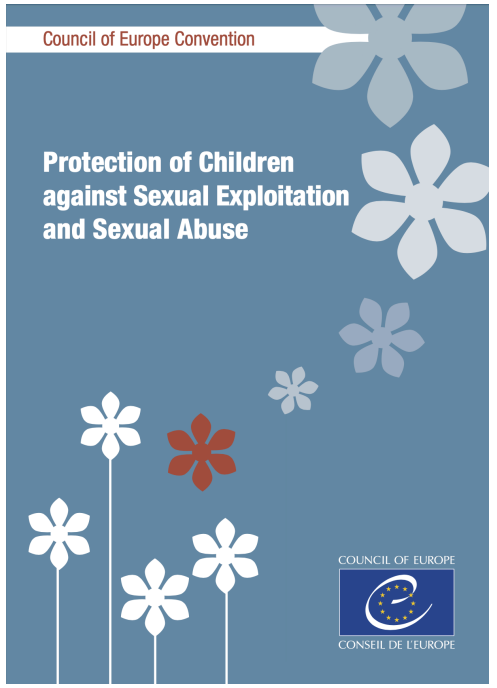
[1]

aka Lanzarote Convention (2007a)

[2]

Child Sexual Abuse and Exploitation (CSAE)

- Child Sexual Exploitation (CSE) [1,2]
  - CSA becomes CSE when a second party benefits monetarily through sexual activity involving a child
    - sexual solicitation or prostitution of a child or adolescent
    - situations where a child or other person is given or promised money or another form of renumeration, payment or consideration in return for the child engaging in sexual activity, even if payment/renumeration is not made



[1]

aka Lanzarote Convention (2007a)

[2]

Child Sexual Abuse and Exploitation (CSAE)

- online abuse
  - no definition in international law [1]
  - UNICEF report defines it as
    - use of the internet, mobile phone or other form of information communication technology to bully, threaten, harass, groom, sexually abuse or sexually exploit a child [1]

[1]

Child Sexual Abuse and Exploitation (CSAE)

- online abuse
  - no definition in international law [1]
  - UNICEF report defines it as
    - use of the internet, mobile phone or other form of information communication technology to bully, threaten, harass, groom, sexually abuse or sexually exploit a child [1]

[1]

Child Sexual Abuse and Exploitation (CSAE) – Examples

- online CSA
  - (cyber) grooming
    - deliberate preparation of a child for sexual abuse or exploitation, motivated by the desire to use the child for sexual gratification [1]

[1]

Child Sexual Abuse and Exploitation (CSAE) – Examples

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Video Live

Three-year-olds groomed online, charity warns

23 April 2024

Share Save

Chris Vallance
Technology reporter

Süddeutsche Zeitung

s | Bundestagswahl USA Klimawandel | Politik Wirtschaft Meinung Panorama Sport Mü

ANZEIGE

Kriminalität › Prozess - Cyber-Grooming von Mädchen: Der Fremde aus dem Internet

Prozess

Cyber-Grooming von Mädchen: Der Fremde aus dem Internet

26. Januar 2023, 17:56 Uhr | Lesezeit: 2 Min.

Child Sexual Abuse and Exploitation (CSAE) – Examples

- online CSA
  - (cyber) grooming
    - deliberate preparation of a child for sexual abuse or exploitation, motivated by the desire to use the child for sexual gratification [1]
  - pressurized sexting
    - forced exchange of user generated sexual imagery or sexual texts via cell phone and other electronic devices [partially 1]

[1]

Child Sexual Abuse and Exploitation (CSAE) – Examples

- online CSA
  - (cyber) grooming
    - deliberate preparation of a child for sexual abuse or exploitation, motivated by the desire to use the child for sexual gratification [1]
  - pressurized sexting
    - forced exchange of user generated sexual imagery or sexual texts via cell phone and other electronic devices [partially 1]
- online CSE
  - sextortion (= sexual extortion of children)
    - blackmailing of a person with intimate images of that person to extort sexual favors, money, or other benefits from them under the threat of sharing the image beyond the previously given consent [1]

[1]

Child Sexual Abuse and Exploitation (CSAE)

The New York Times

How to Protect Your Children From Online Sexual Predators

Share full article

Parental Warning:

Parents should carefully monitor their children's activity on these popular social media apps and games.

Whisper, Skout, Grindr, Omegle, Tinder, Chat Avenue, Chat Roulette, Wishbone, Live.ly, Musical.ly, Paltalk, Yubo, KiK, Hot or Not, Down, Tumblr, Games: Fortnite, Minecraft, Discord

A police notice to parents on the dangers minors face online. New Jersey State Police

By Michael H. Keller

The New York Times

'Chelsea' Asked for Nude Pictures. Then the Sextortion Began.

Young men are being tricked into sending naked pictures to scammers pretending to be women — who then demand money. The consequences can be devastating.

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Video Live

'I thought my life was over': Escaping the sextortion scammers

12 September 2024

Share Save

Jayne McCubbin
BBC News

Süddeutsche Zeitung

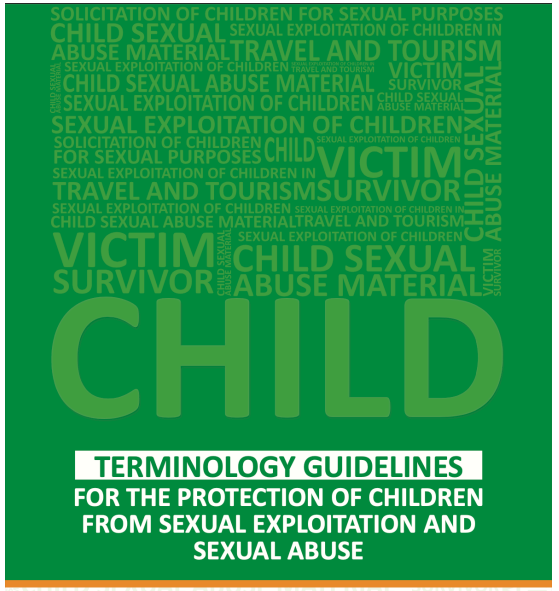
Sextortion

"Dann wirst du tun, was immer ich verlange"

intimate imagery

Child Sexual Abuse Material (CSAM)

- CSAM
  - any representation of a child depicting acts of sexual abuse and/or focusing on the genitalia of the child [1]
  - generation [1]
    - self-generated by the minor
    - perpetrator generated
    - created virtually
  - problematic through creation but also revictimization through (re-)sharing

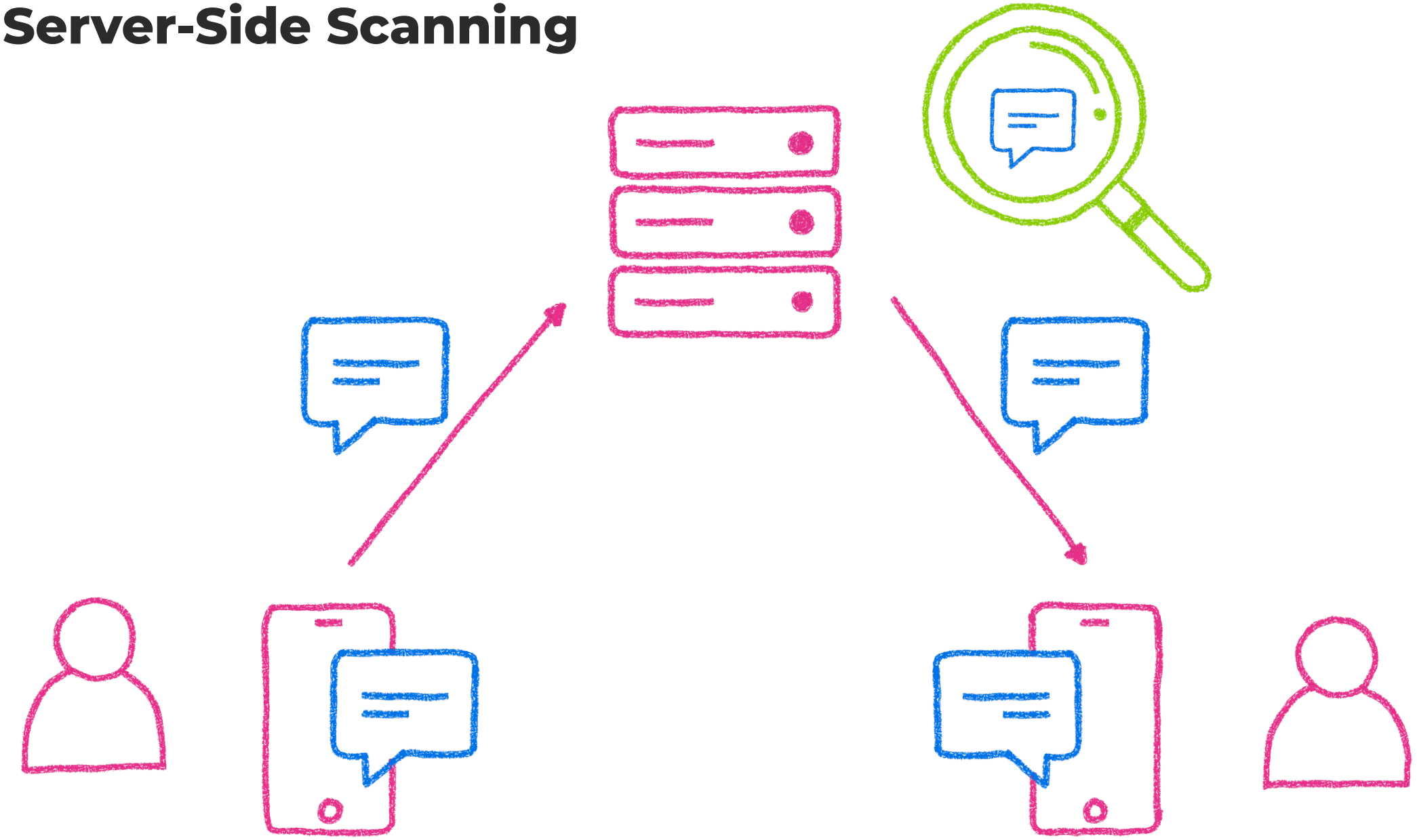


[1]

> 99.5 %

were classified as child sexual abuse material (CSAM)

Server-Side Scanning



Server-Side Scanning

The New York Times

A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal.

Google has an automated tool to detect abusive images of children. But the system can get it wrong, and the consequences are serious.

The Guardian

Eur

Technology

Google refuses to reinstate man's account after he took medical images of son's groin

Experts say case highlights dangers of automated detection of child sexual abuse images

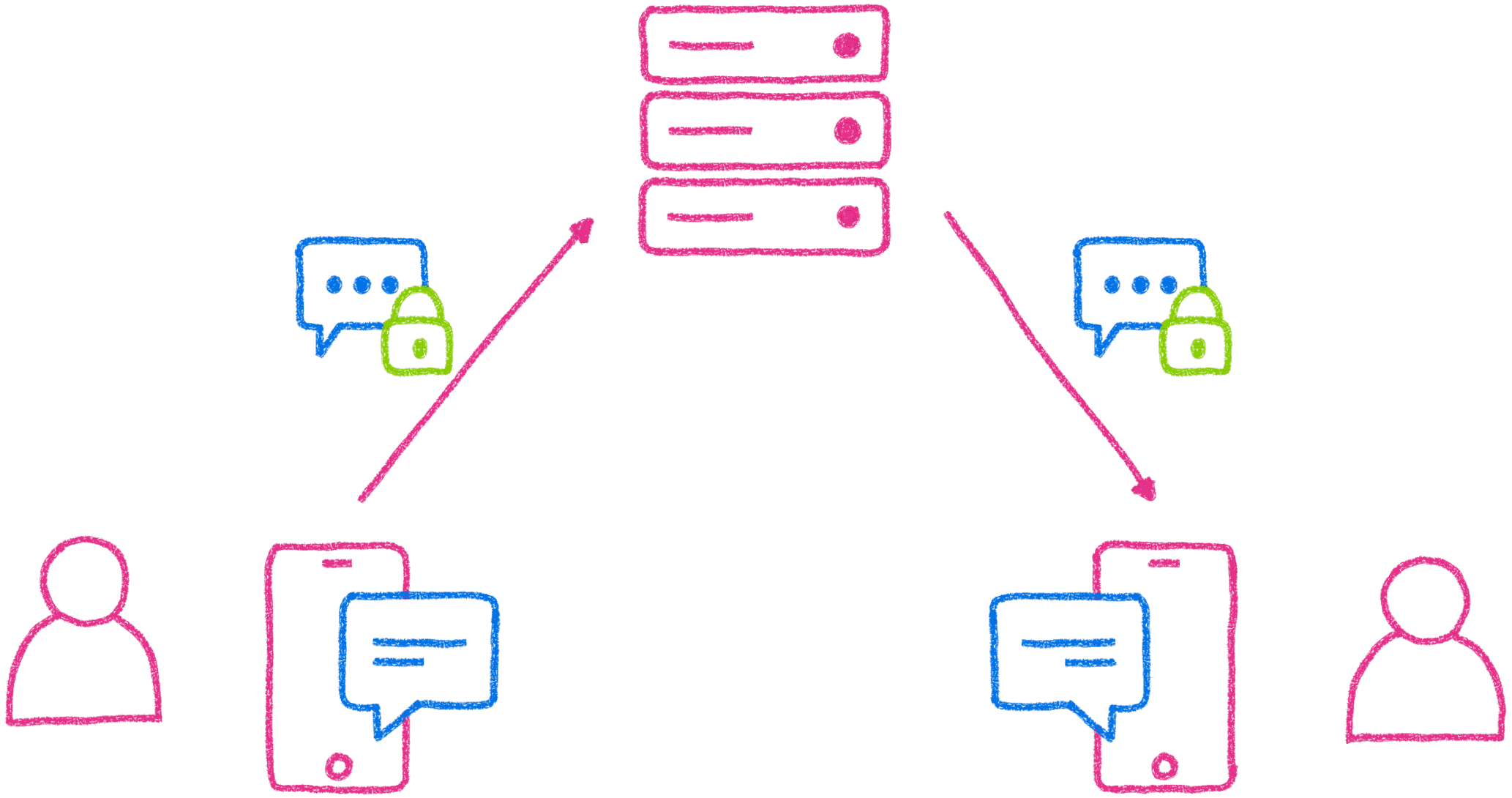
Johana Bhuiyan

Tue 23 Aug 2022 01.32 CEST

Share

Tech companies like Google have access to a vast trove of data – but no context for it, says an ACLU technologist. Photograph: Avishek Das/Sopa Images/Rex/Shutterstock

End-To-End Encryption



End-To-End Encryption

The New York Times

WhatsApp Introduces End-to-End Encryption

By Mike Isaac

April 5, 2016

The New York Times

Meta Plans to Add Encryption to Messenger, Stoking a Privacy Debate

The move is part of an effort to make the app more like WhatsApp and iMessage. Law enforcement authorities say the privacy makes it harder to track criminals.

Share full article

"This means that nobody, including Meta, can see what's sent or said, unless you choose to report a message to us," wrote Loredana Crisan, a vice president of Messenger. Godofredo A. Vásquez/Associated Press


By Mike Isaac and Michael H. Keller

Mike Isaac has covered Meta and its messaging services since 2010. Michael H. Keller has written extensively about online safety and messaging apps.

Dec. 6, 2023

Security Research

Overview Blog Bounty Research Device Submit a Report

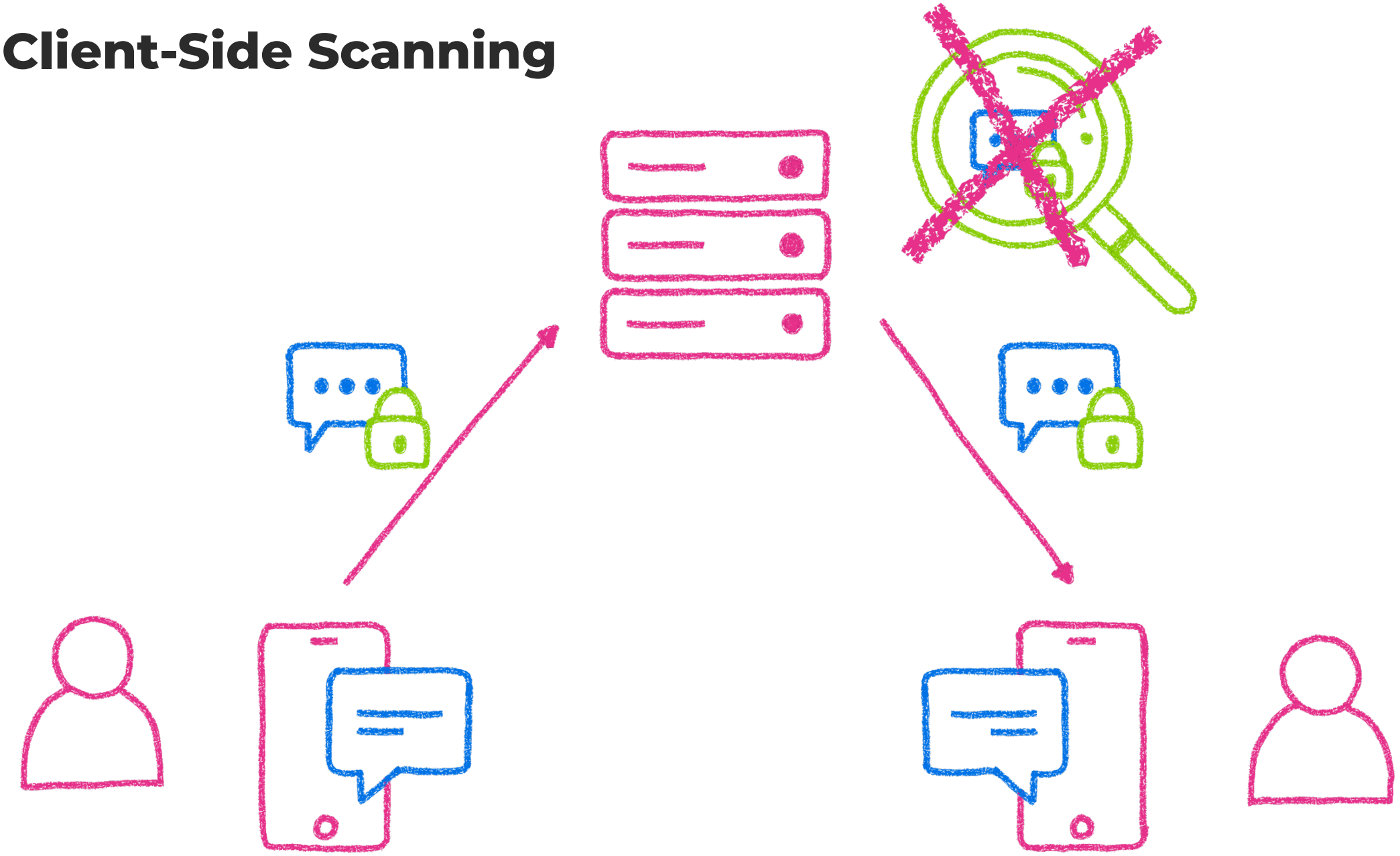
Blog

February 21, 2024

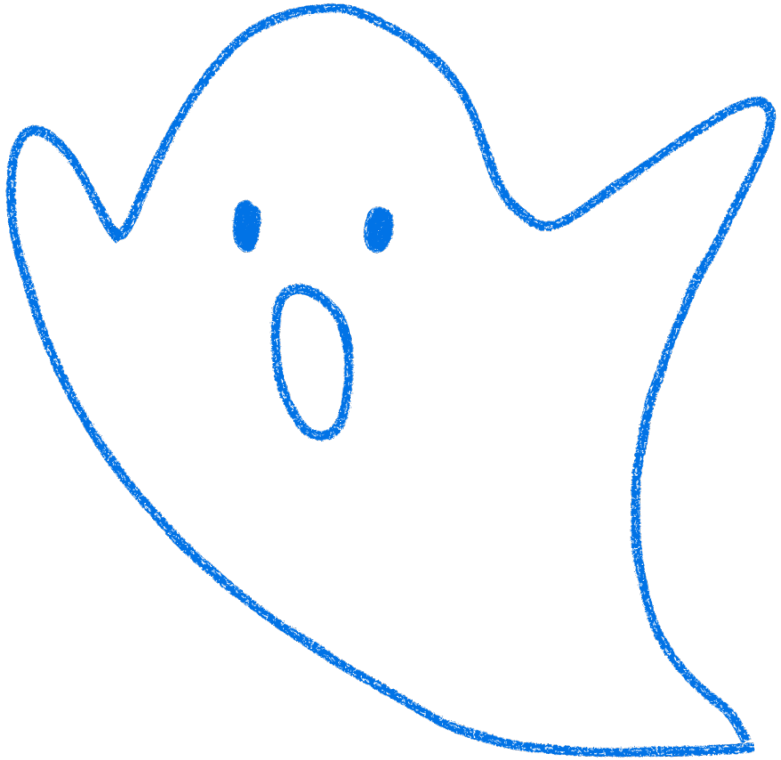
iMessage with PQ3: The new state of the art in quantum-secure messaging at scale

Posted by Apple Security Engineering and Architecture (SEAR)

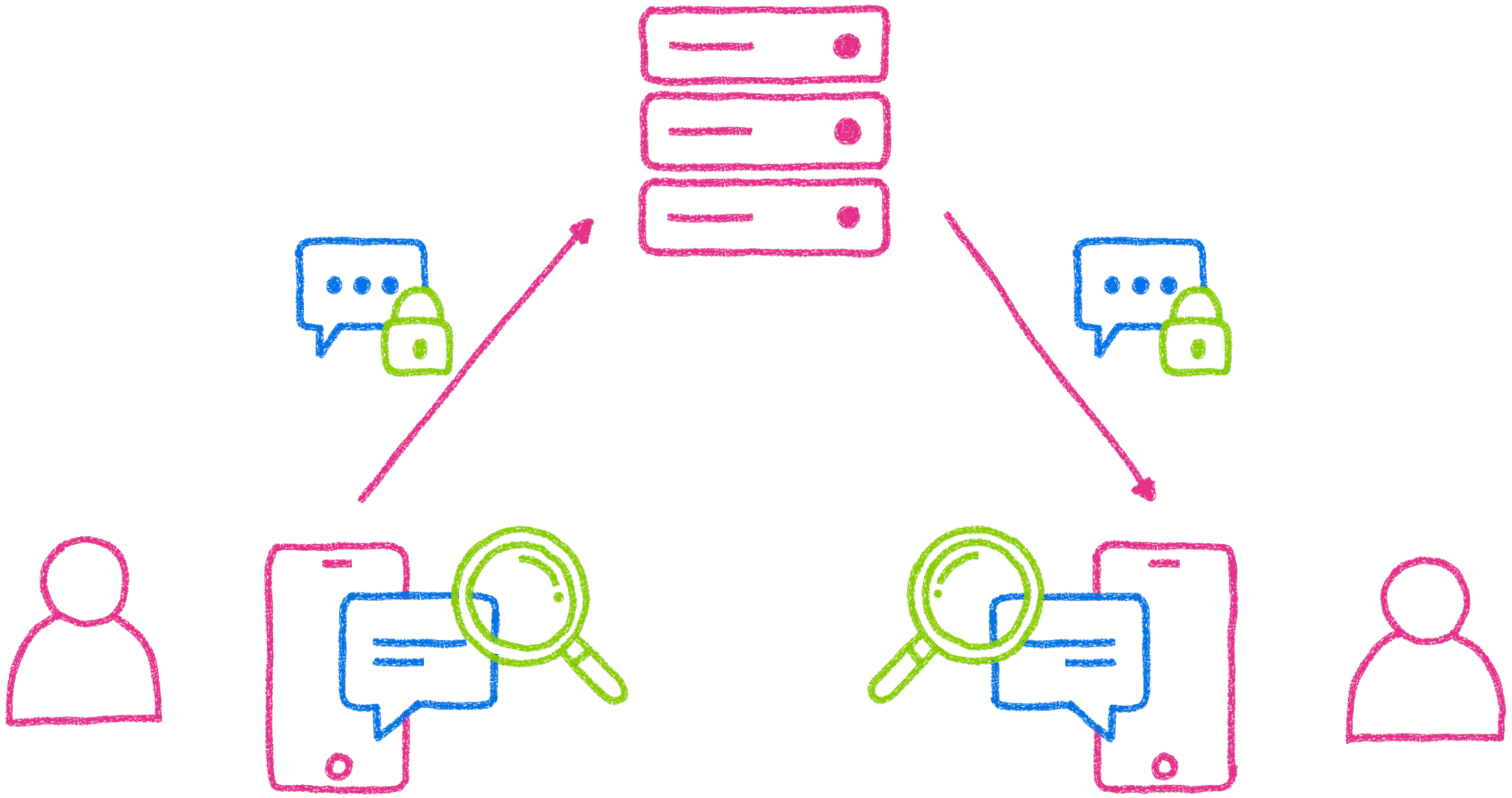
Client-Side Scanning



One Idea: The Ghost Proposal

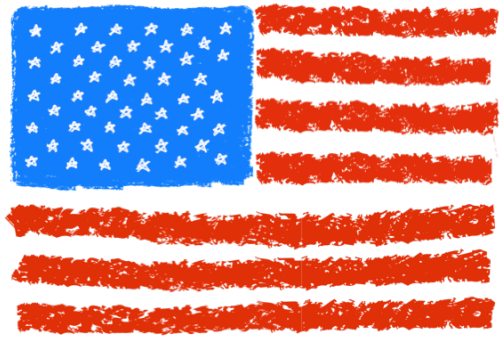


Client-Side Scanning



Proposals

legislative proposals



What is Client-Side Scanning?

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

UNIVERSITÄT DES SAARLANDES

Divyanshu Bhardwaj★

Carolyn Guthoff★

Adrian Dabrowski

Sascha Fahl

Katharina Krombholz

★ both authors contributed equally

Mental Models, Expectations and Implications of Client-Side Scanning: An Interview Study with Experts

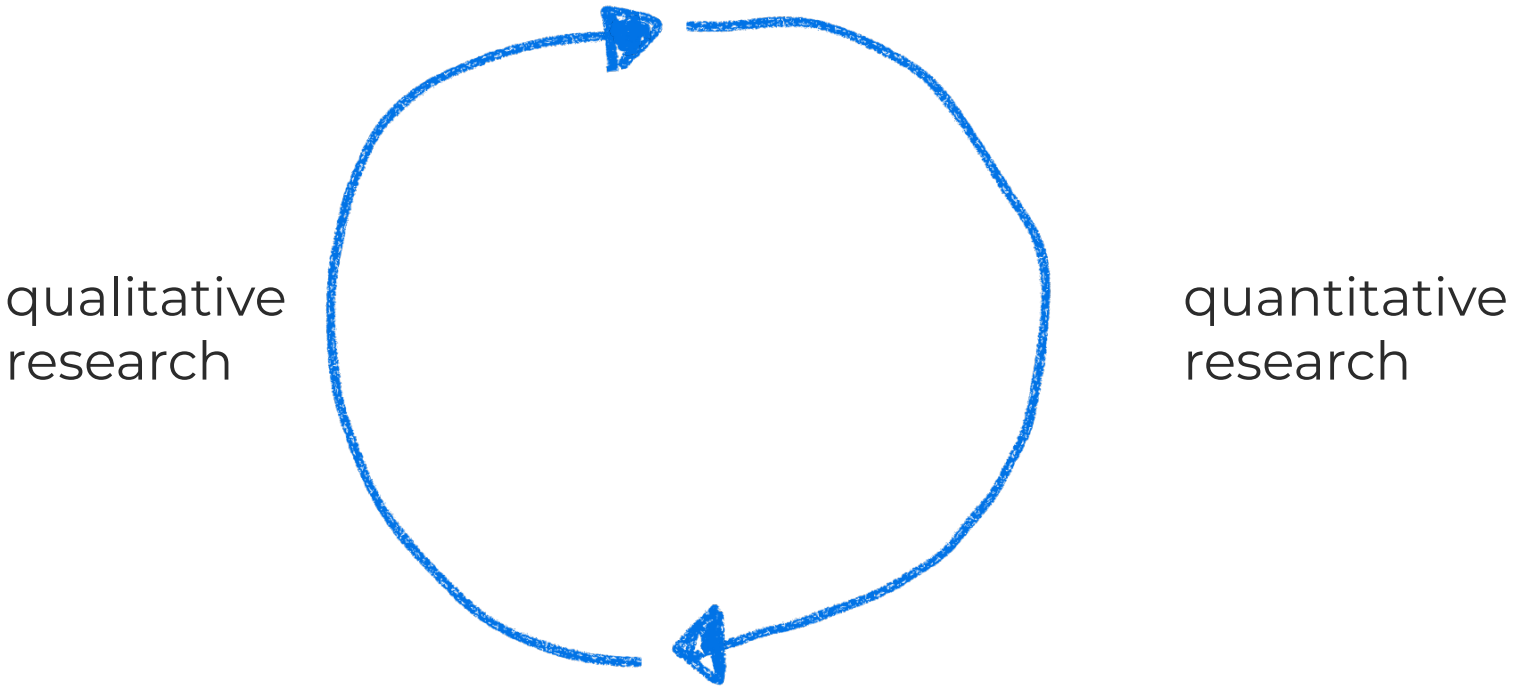
Divyanshu Bhardwaj‡
CISPA Helmholtz Center for Information Security, *and* Saarland University
Germany

Carolyn Guthoff ‡
CISPA Helmholtz Center for Information Security, *and* Saarland University
Germany

Adrian Dabrowski
CISPA Helmholtz Center for Information Security
Germany

Sascha Fahl
CISPA Helmholtz Center for Information Security
Germany

Katharina Krombholz
CISPA Helmholtz Center for Information Security
Germany


qualitative
research
quantitative
research

Methodology

RQ1: mental models

RQ2: expectations

RQ3: implications

Methodology – Key Factors

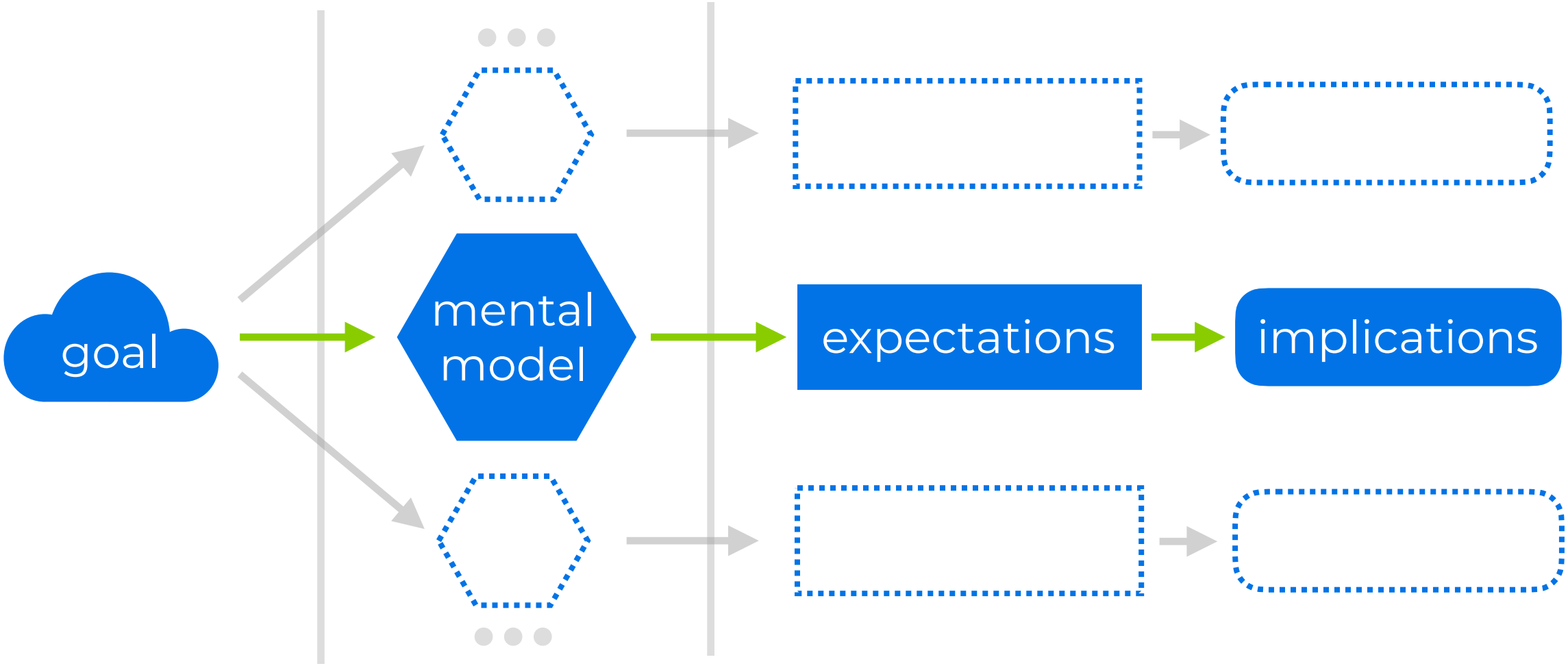
28 participants

25 semi-structured interviews

data analysis

RESULTS

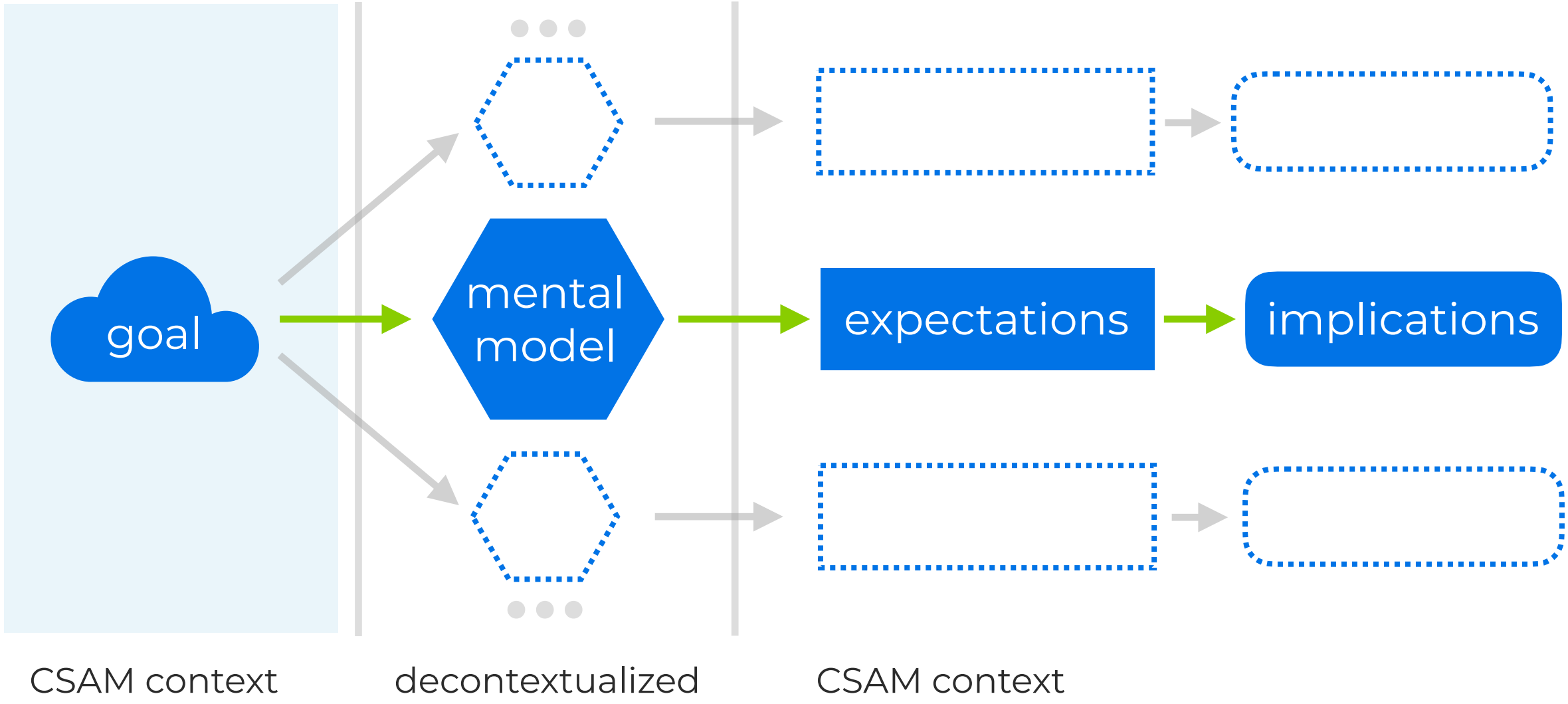
Results – Overview



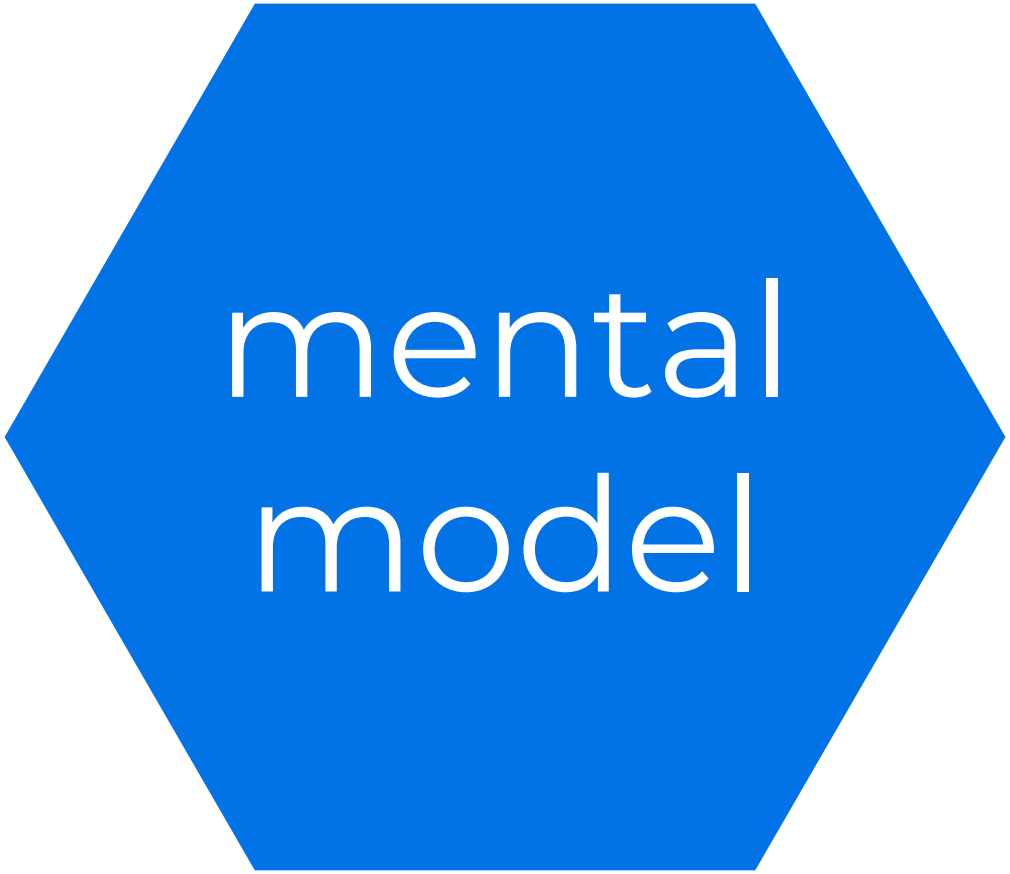
Results – Goal

CSAM context

Results – Overview

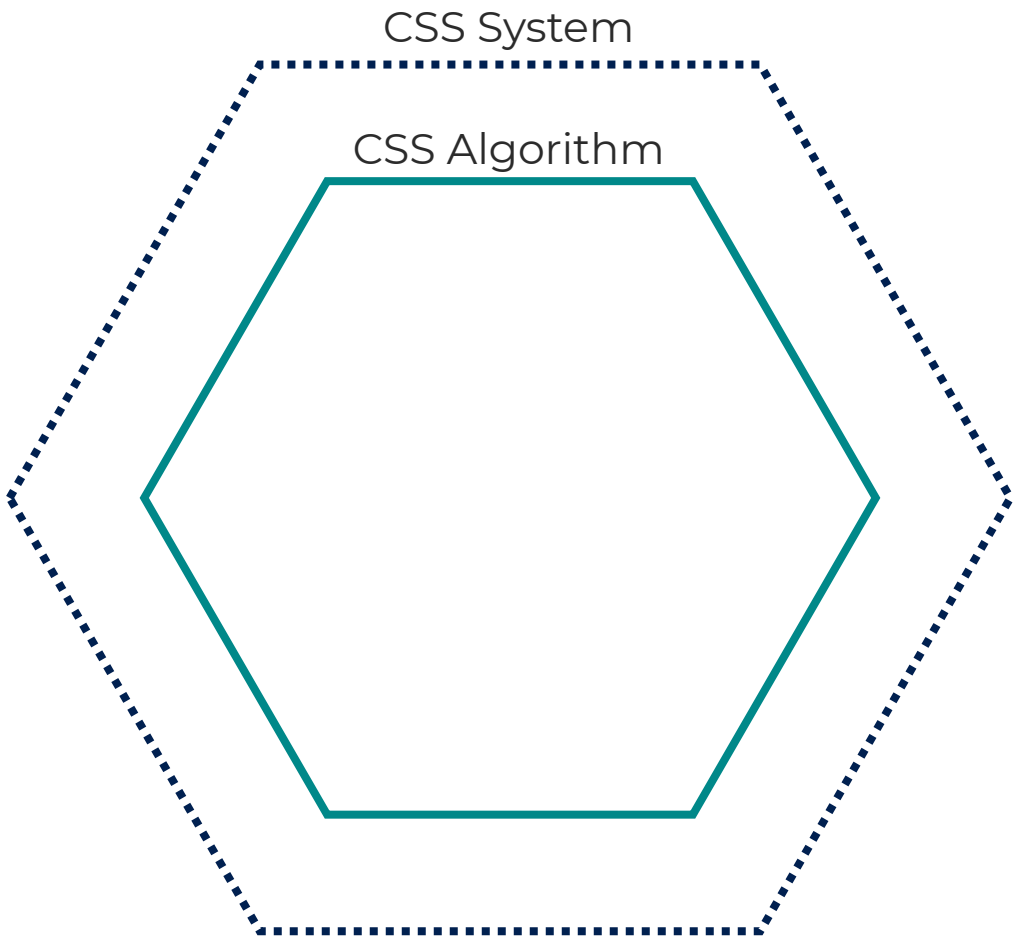


Results – Mental Models



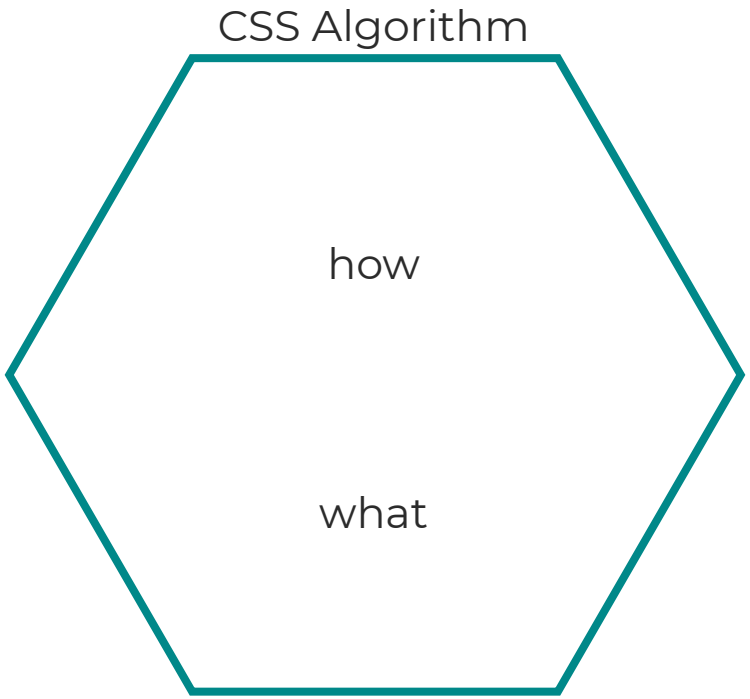
decontextualized

Results – Mental Models



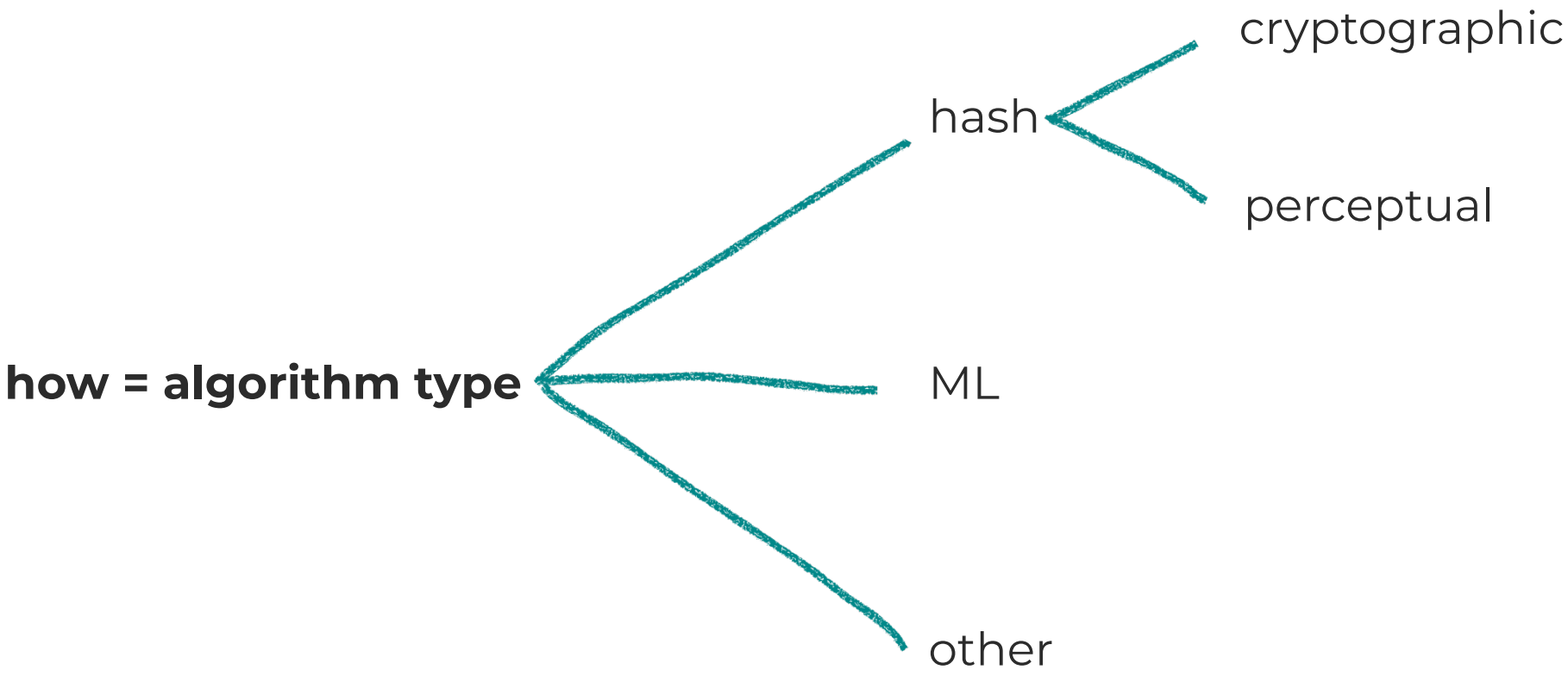
decontextualized

Results – Mental Models



decontextualized

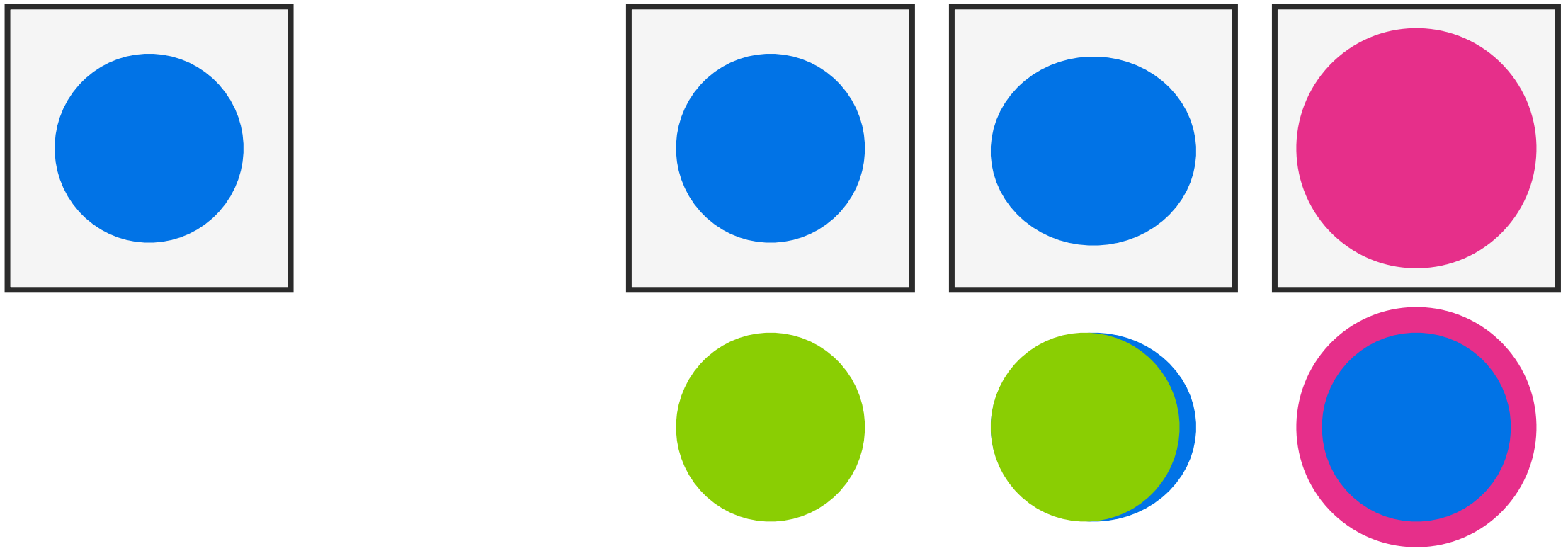
Results – CSS Algorithm – how



Results – CCS Algorithm – how

comparison image

reference images

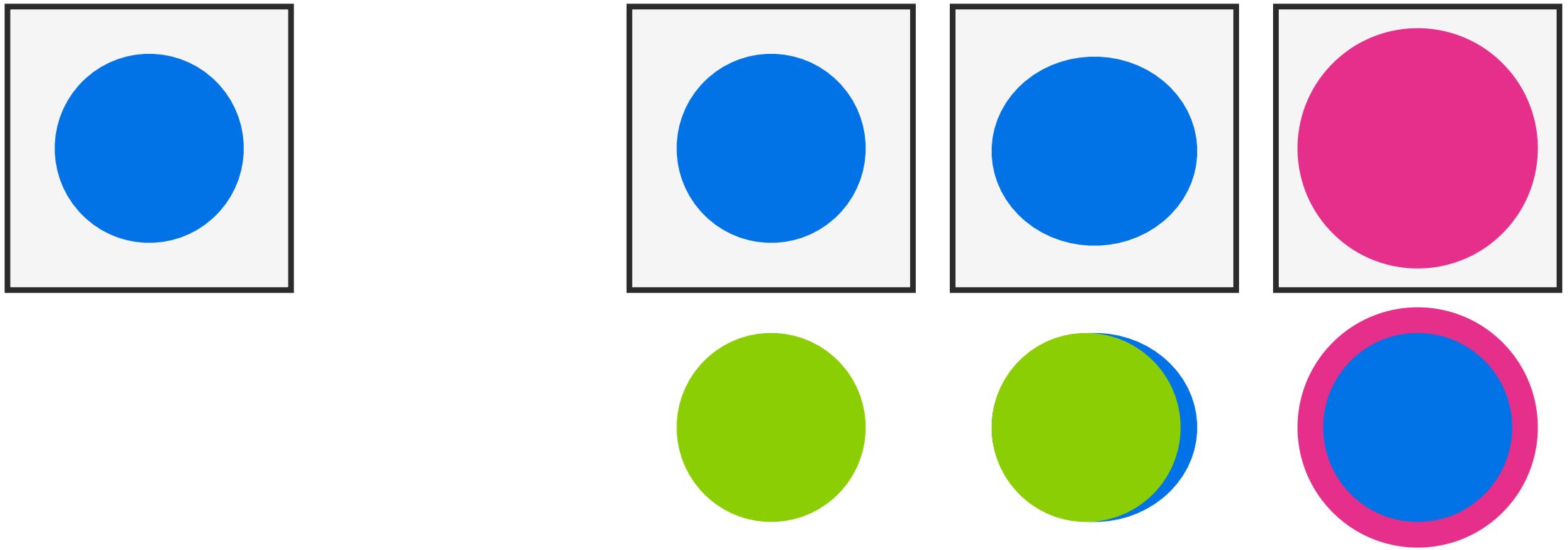


cryptographic hash	✅	❌	❌

Results – CCS Algorithm – how

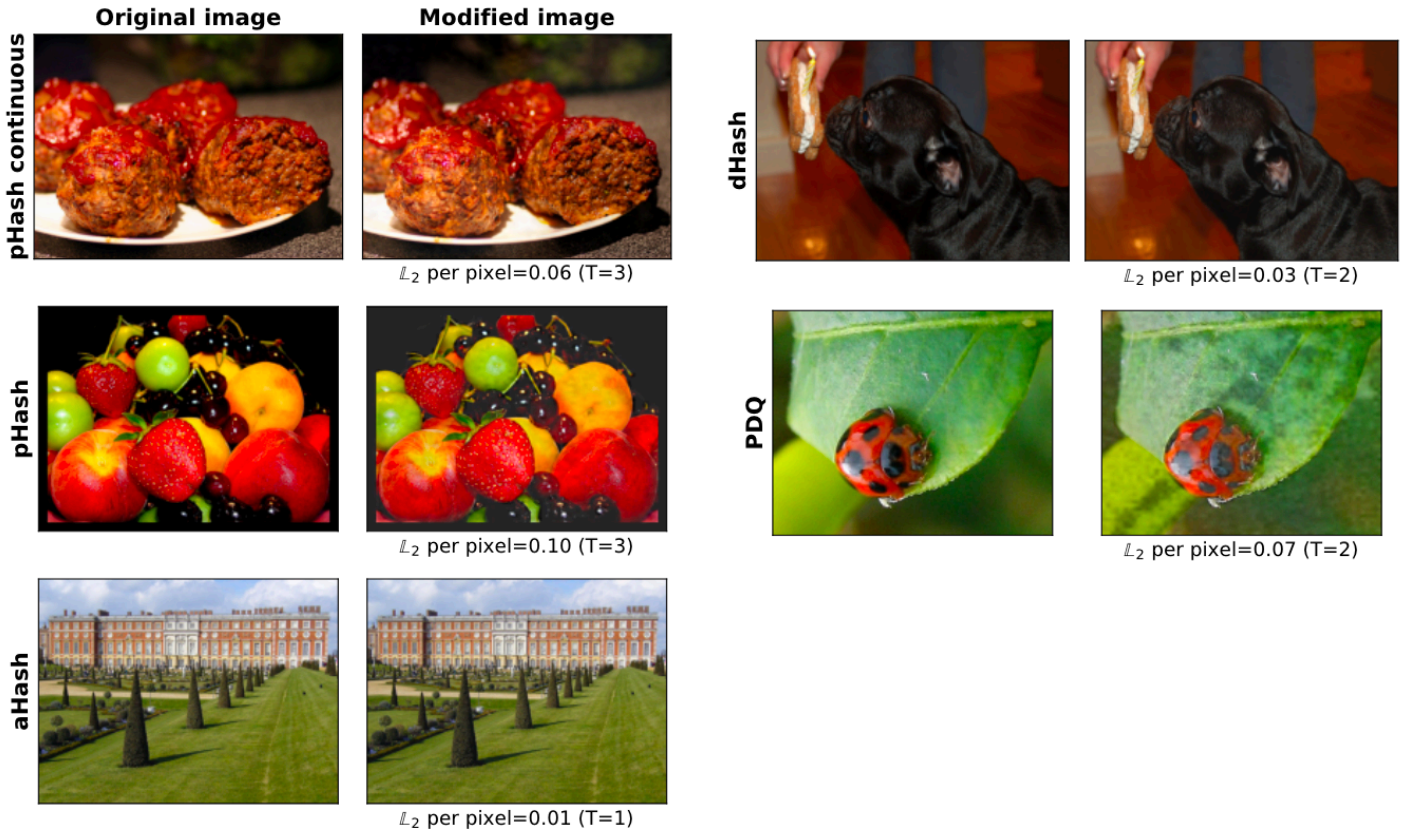
comparison image

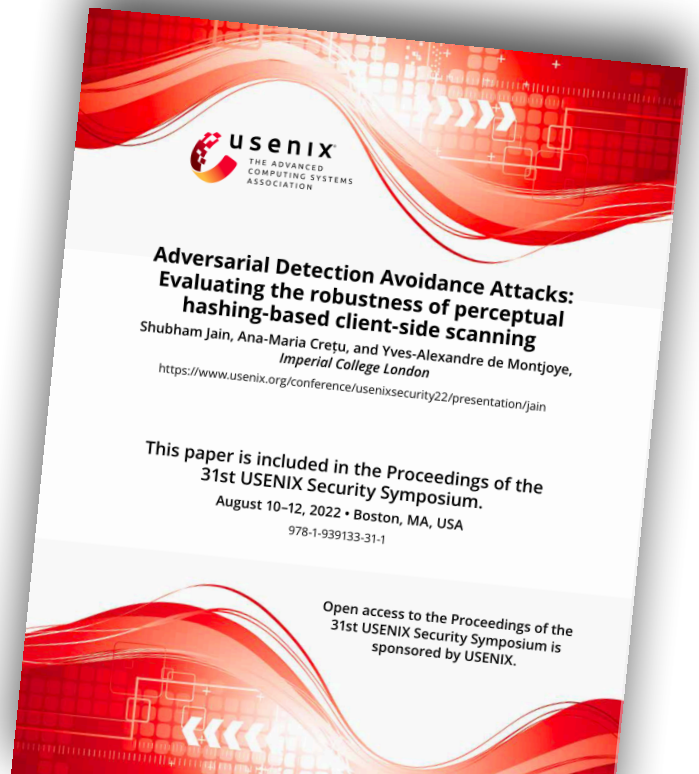
reference images



cryptographic hash	✅	❌	❌
perceptual hash	✅	❓✅	❌

Perceptual Hashes



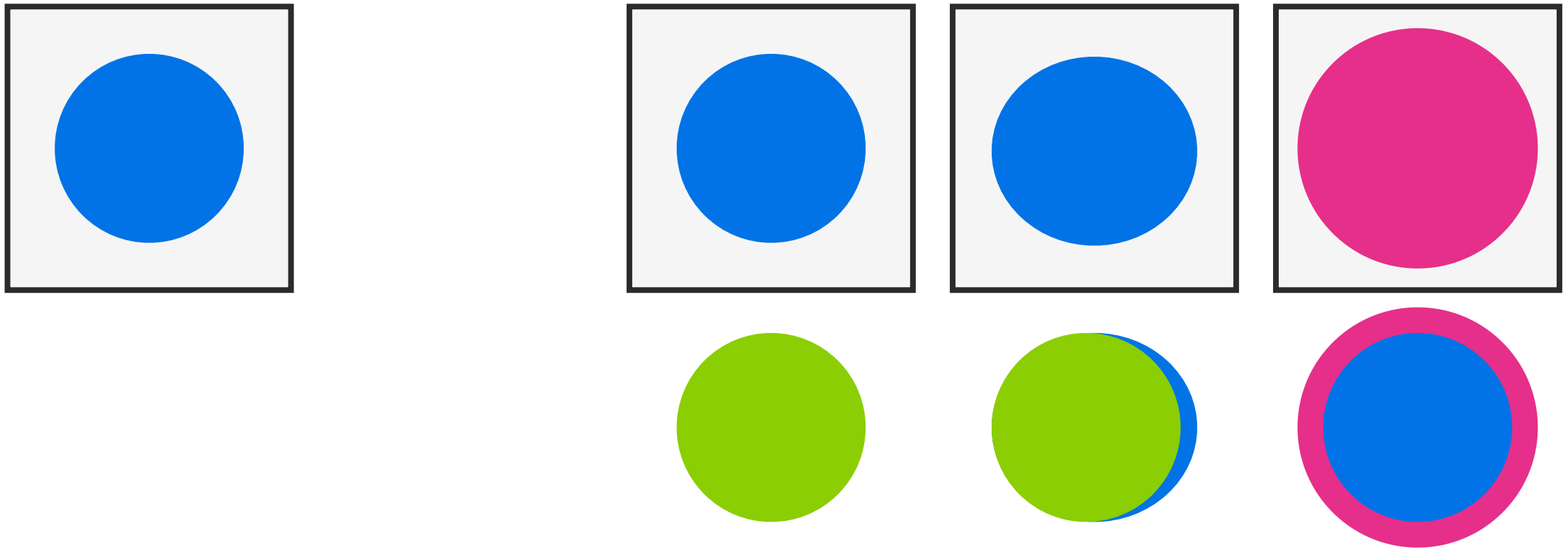


Jain, Cretu and de Montjoye - Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning (USENIX Security '22)

Results – CCS Algorithm – how

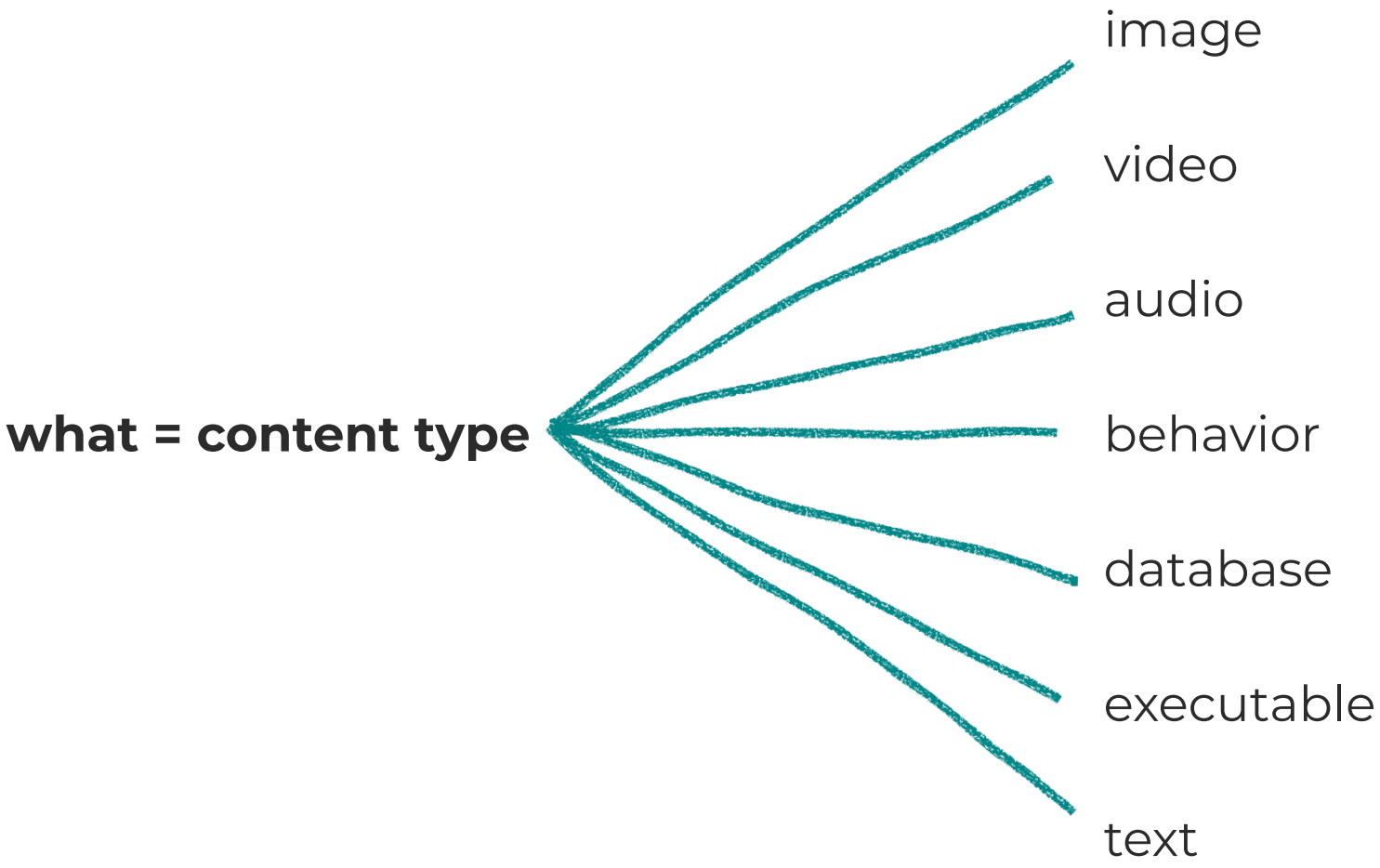
comparison image

reference images

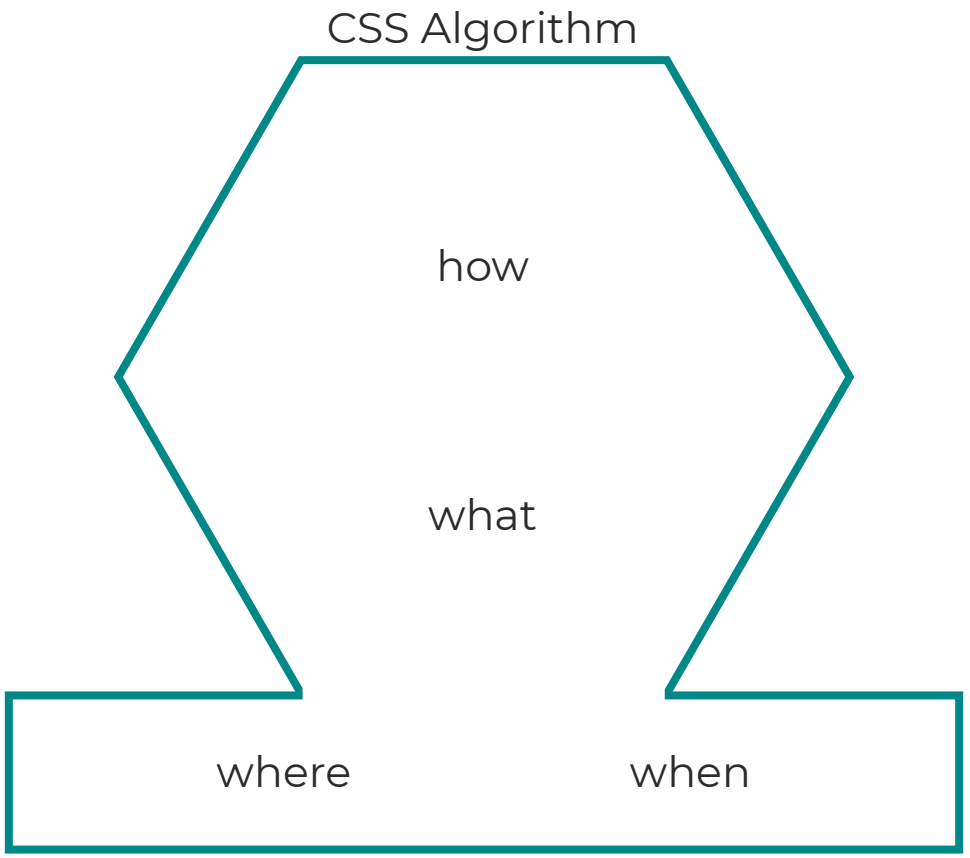


cryptographic hash	✅	❌	❌
perceptual hash	✅	❓✅	❌
ML	❓✅	❓	❓

Results – CSS Algorithm – what

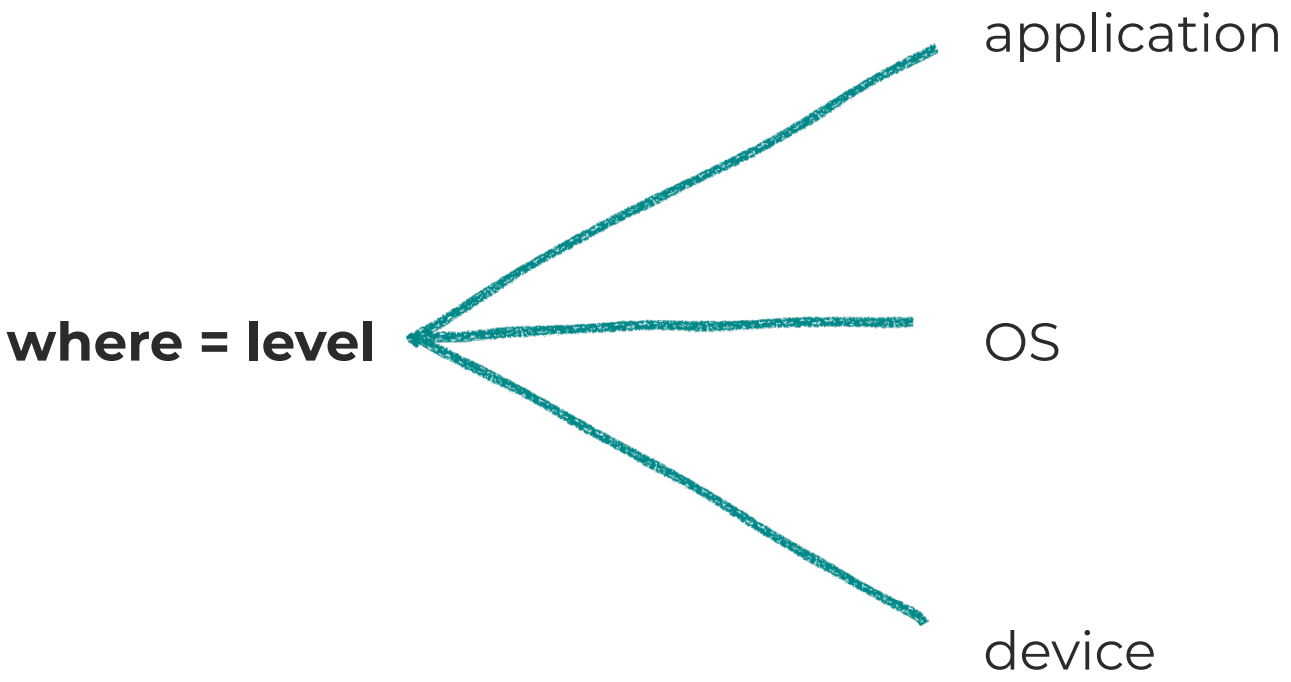


Results – Mental Models

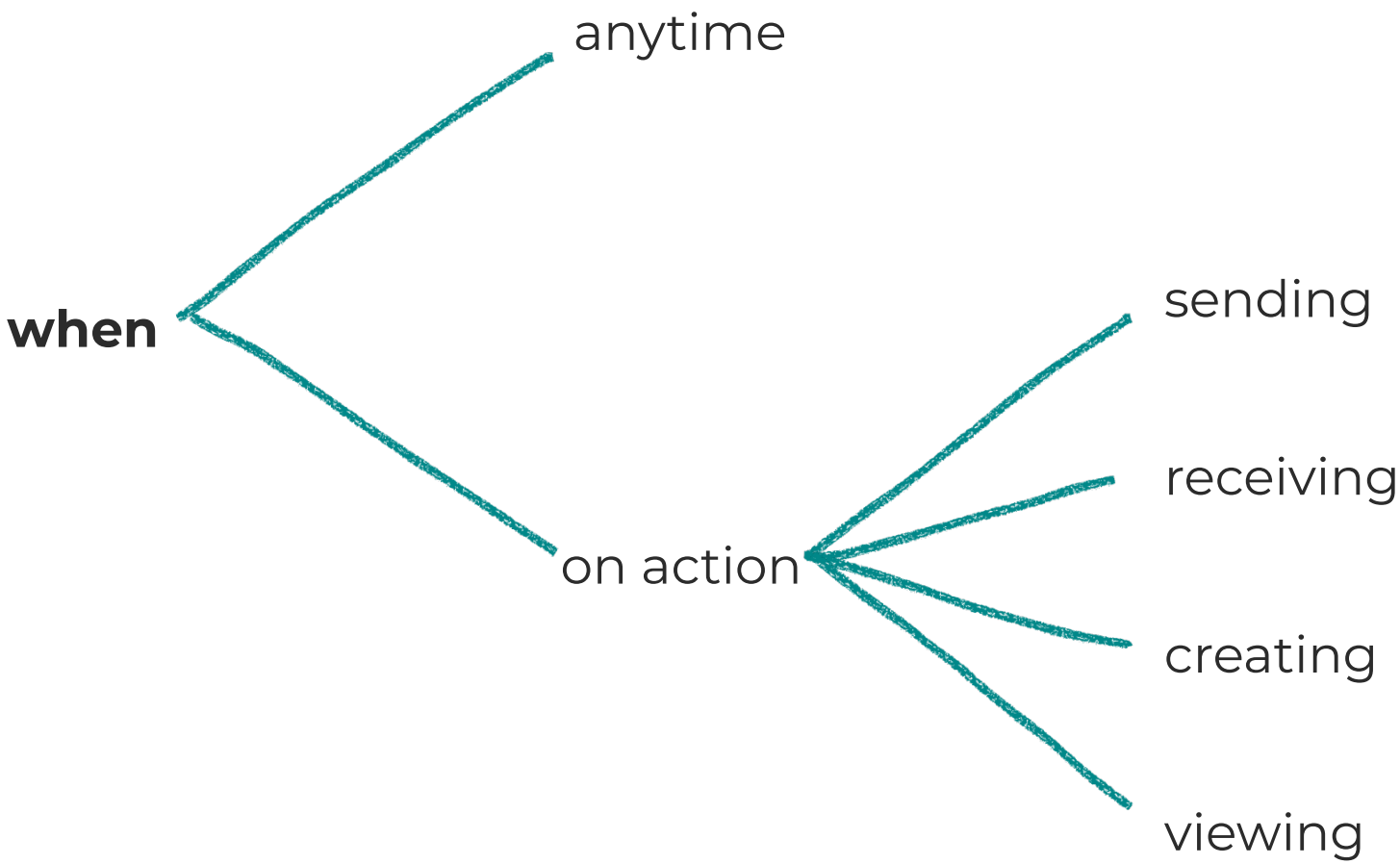


decontextualized

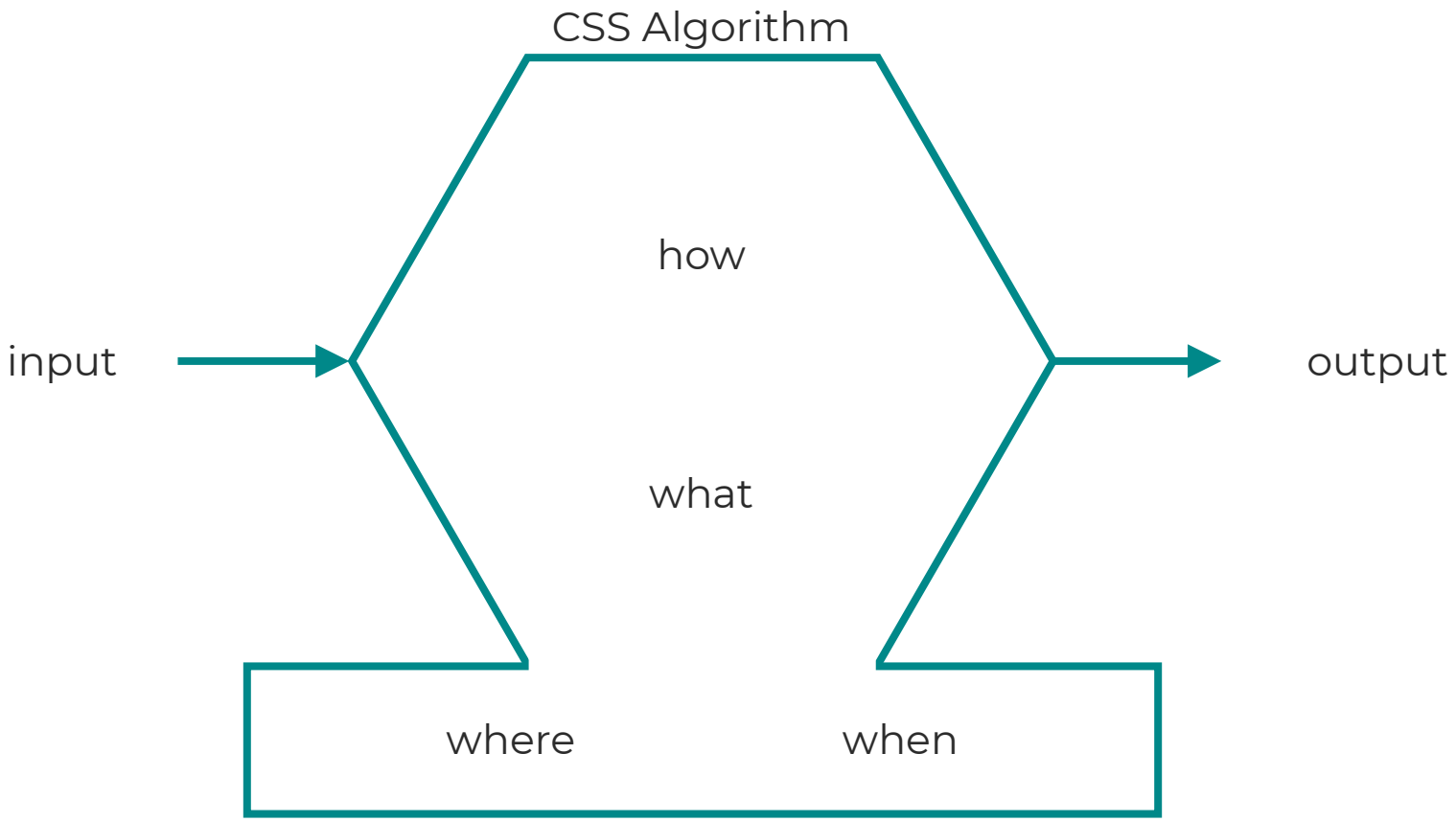
Results – CSS Algorithm – where



Results – CSS Algorithm – when

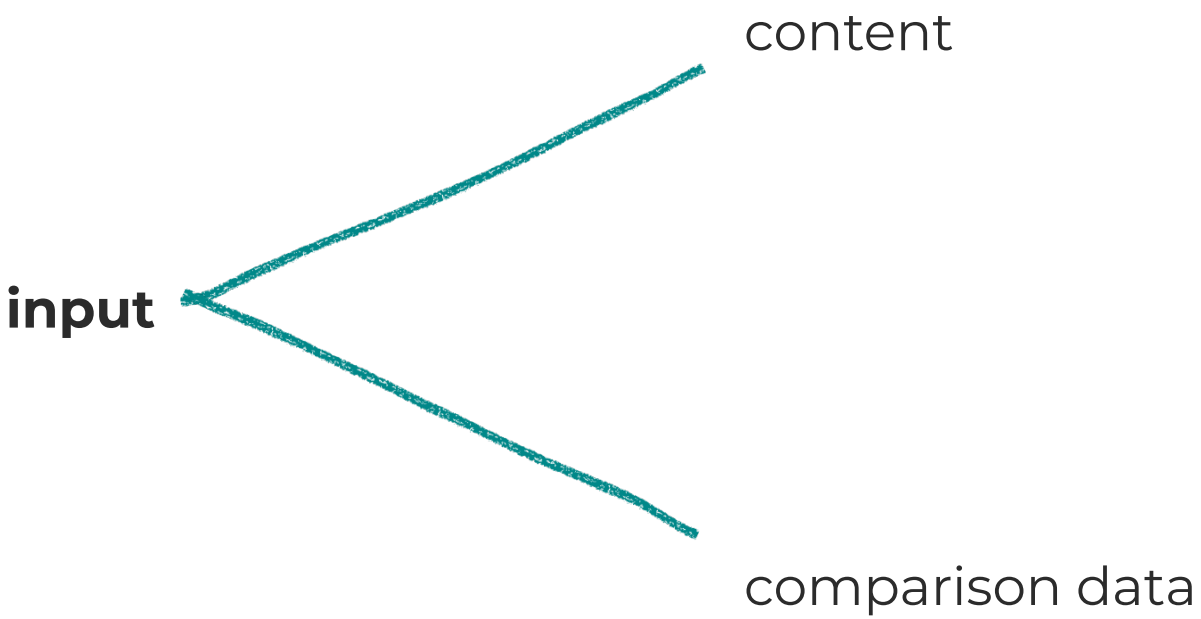


Results – Mental Models

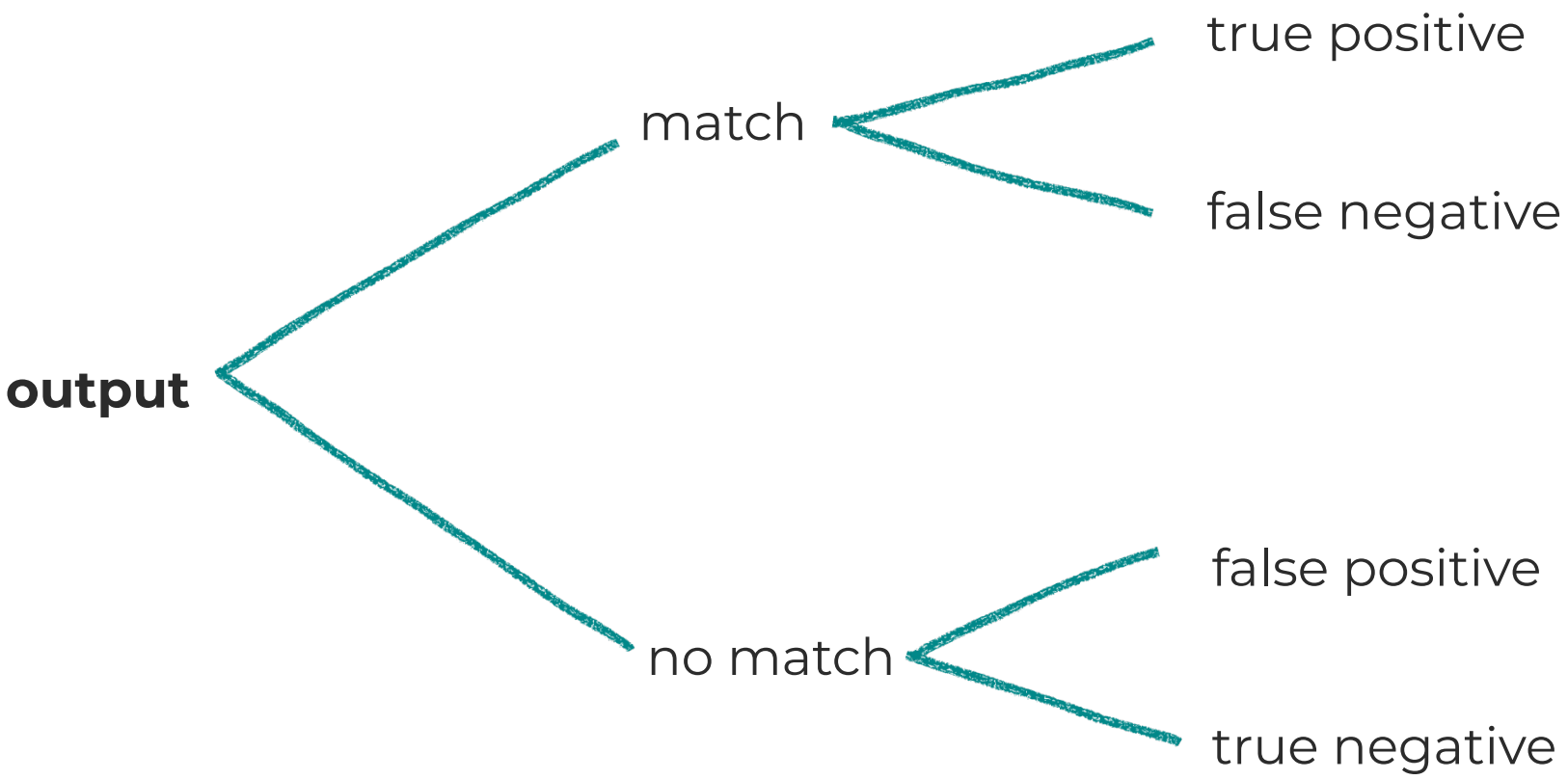


decontextualized

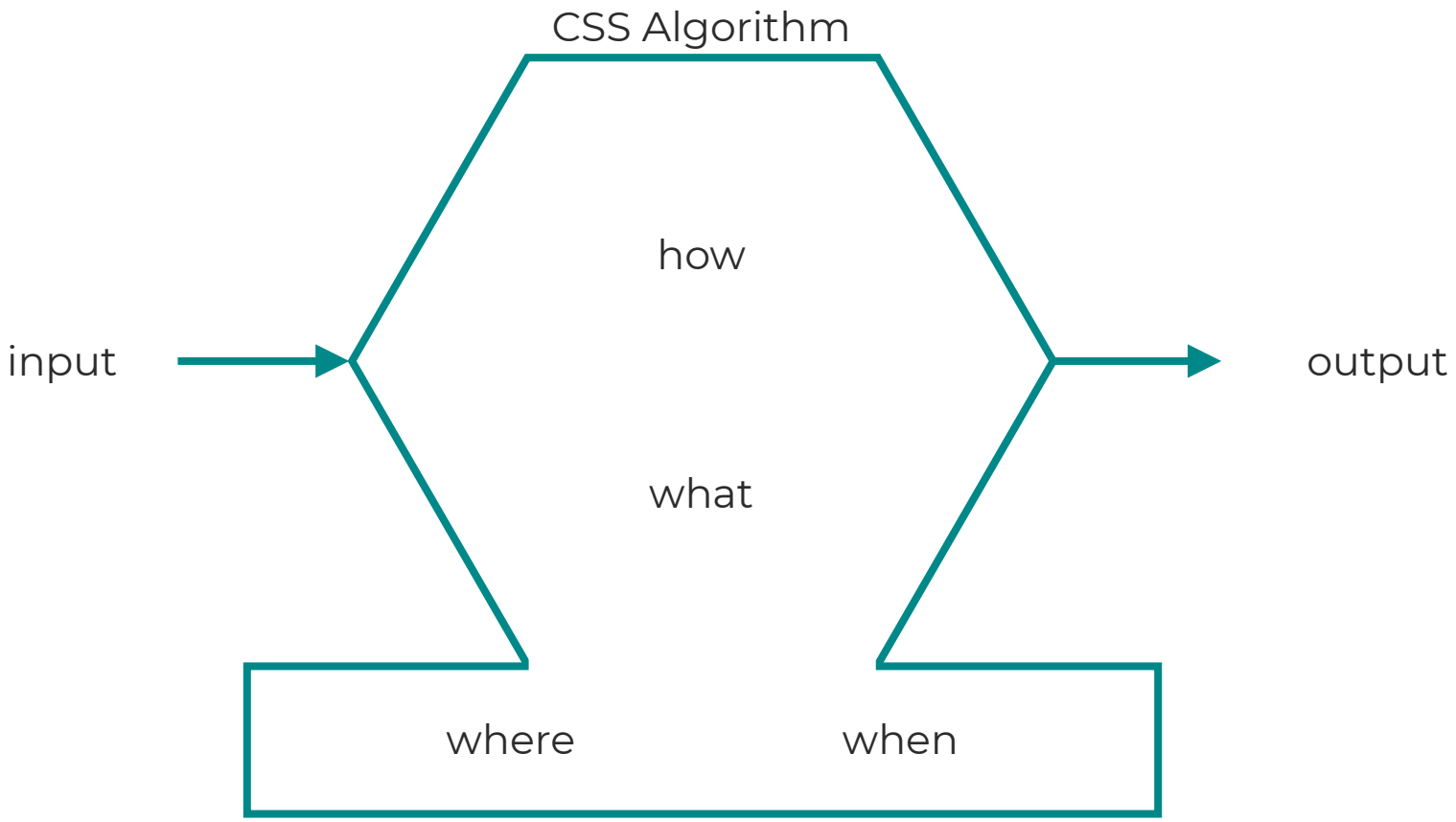
Results – CSS Algorithm – input



Results – CSS Algorithm – output

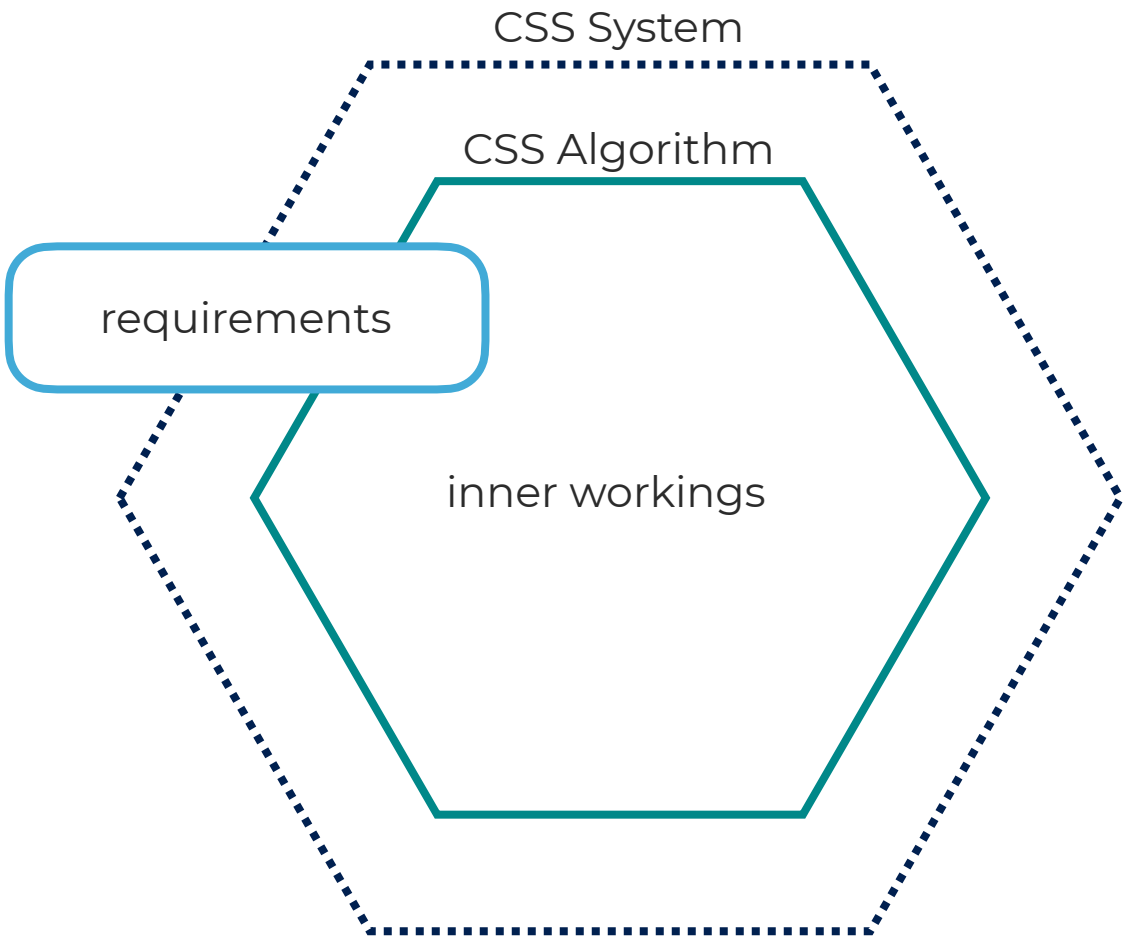


Results – Mental Models



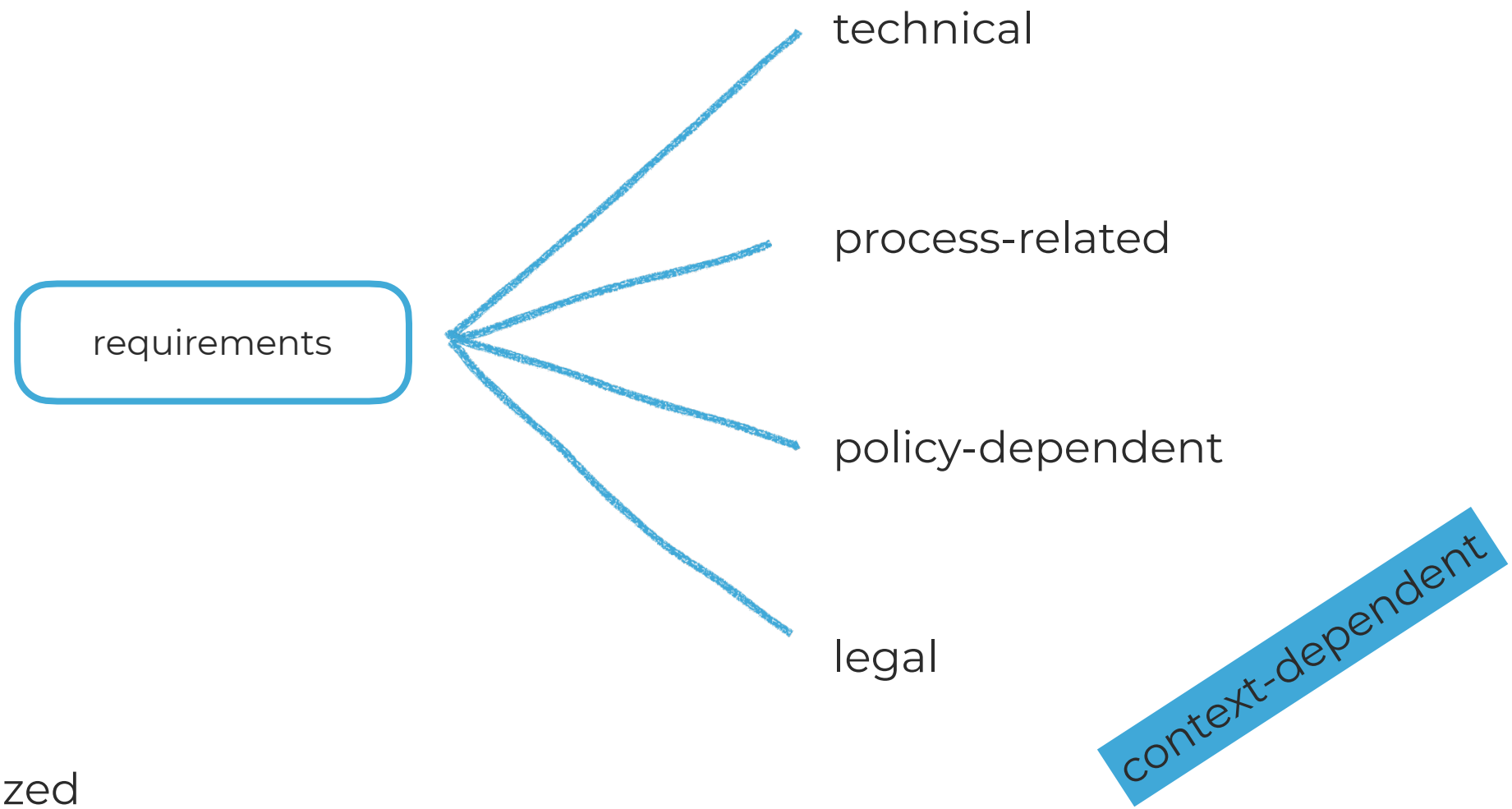
decontextualized

Results – Mental Models



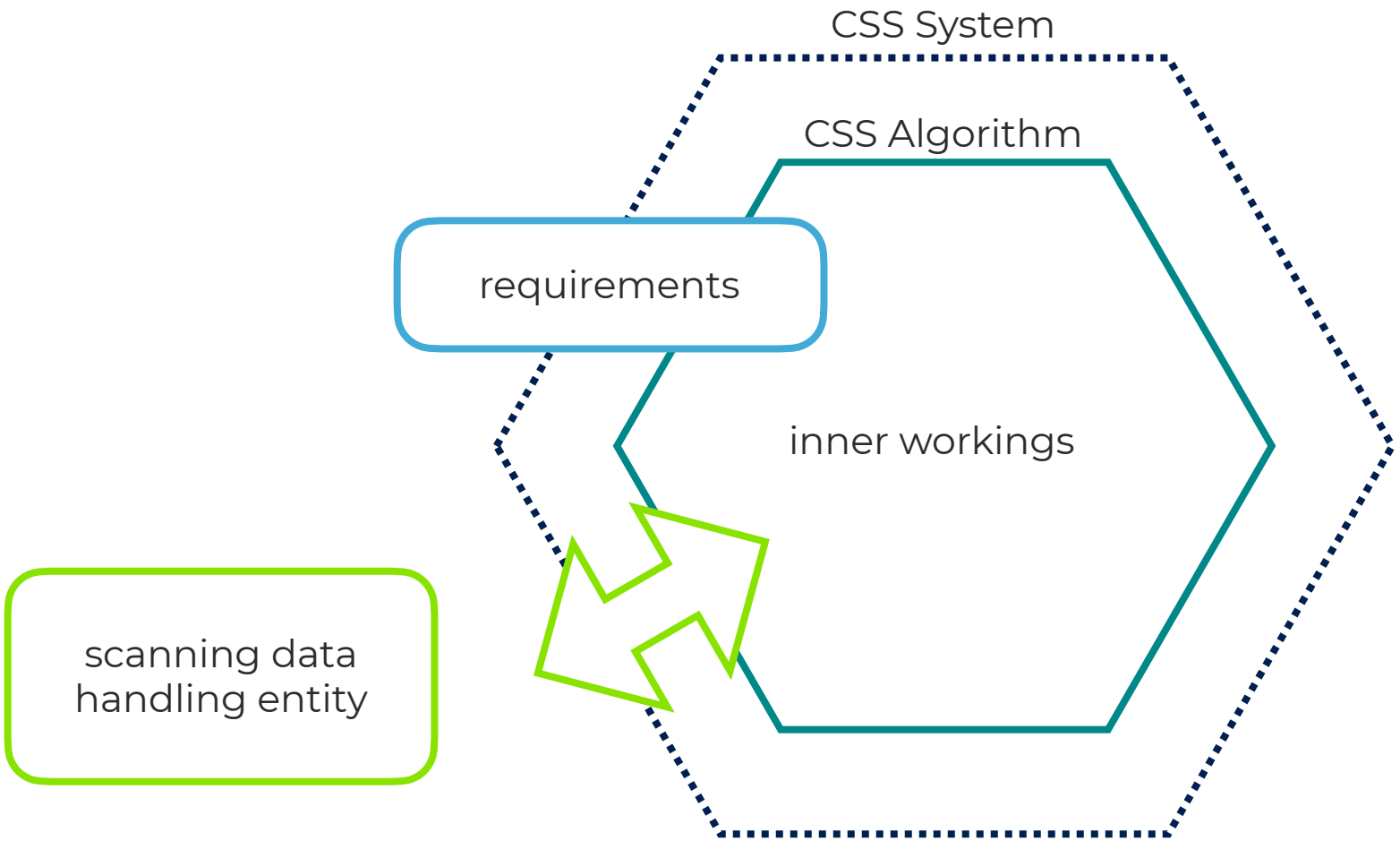
decontextualized

Results – Mental Models



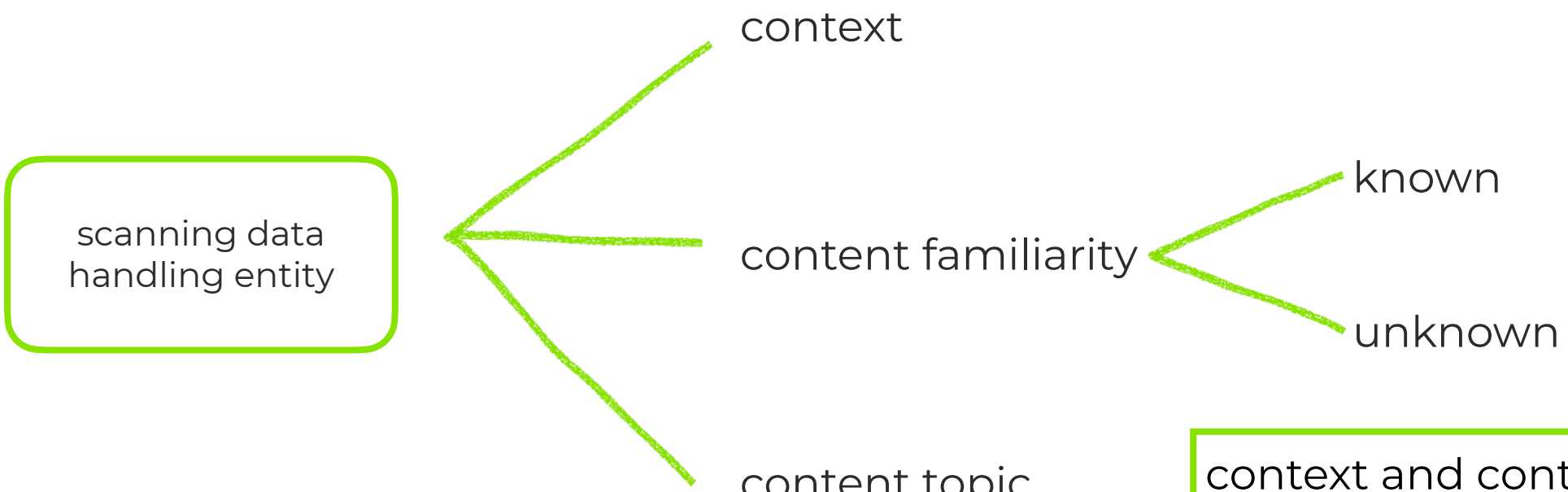
decontextualized

Results – Mental Models



decontextualized

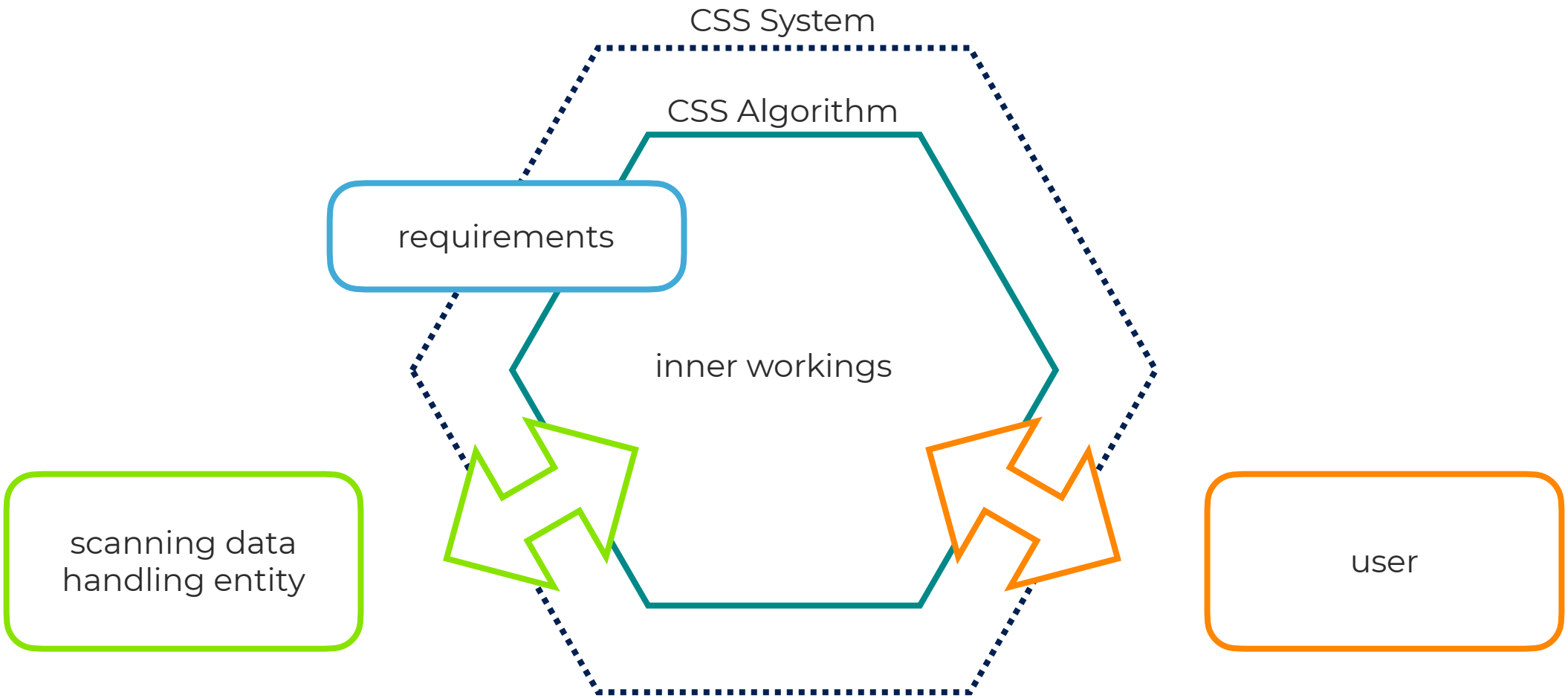
Results – Mental Models



decontextualized

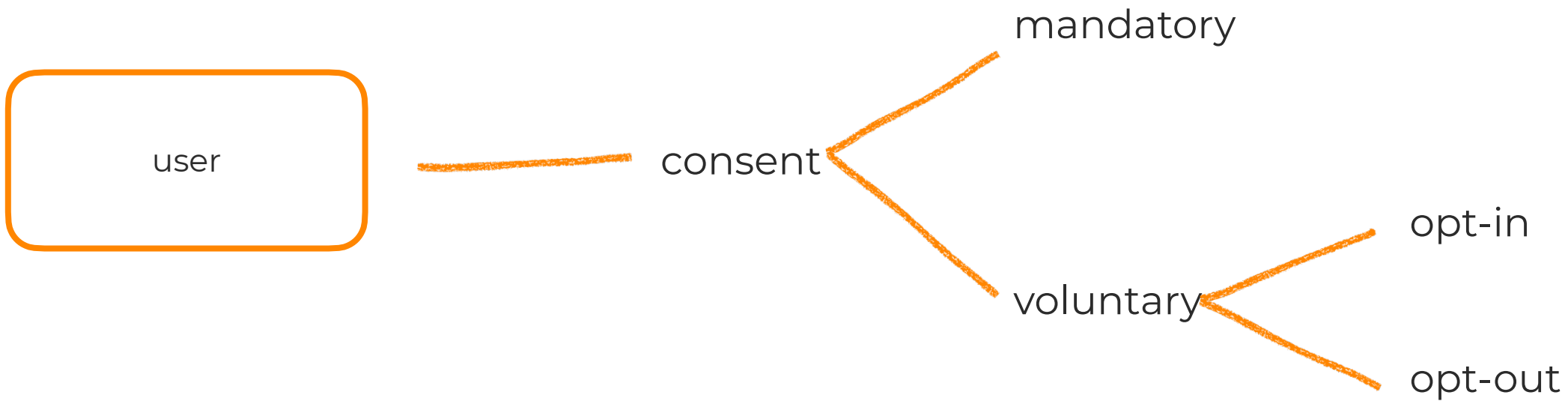
context and content topic influence content type
content familiarity influences the algorithm type

Results – Mental Models



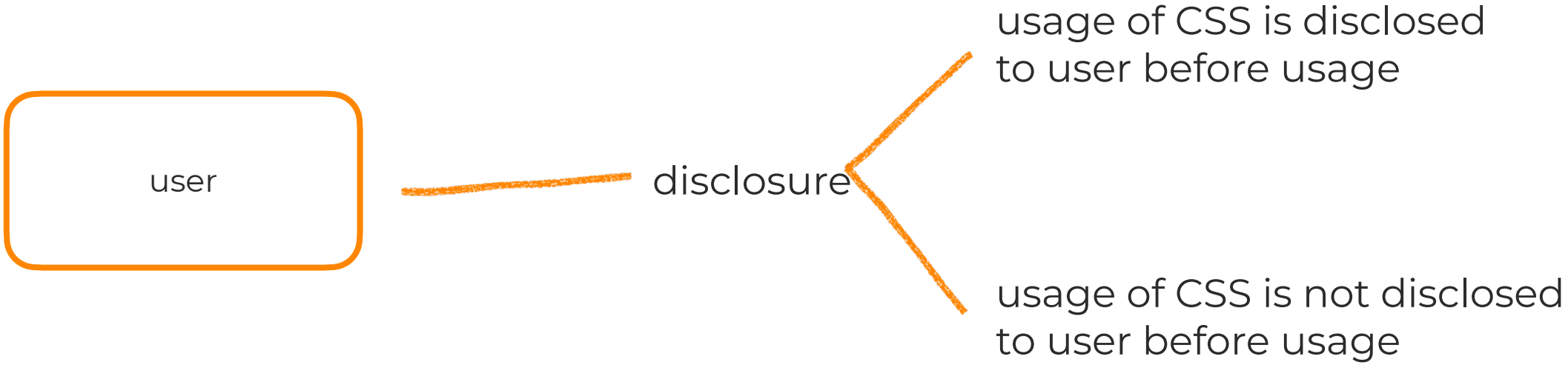
decontextualized

Results – Mental Models



decontextualized

Results – Mental Models



decontextualized

Results – Mental Models

decontextualized

Results – Mental Models

decontextualized

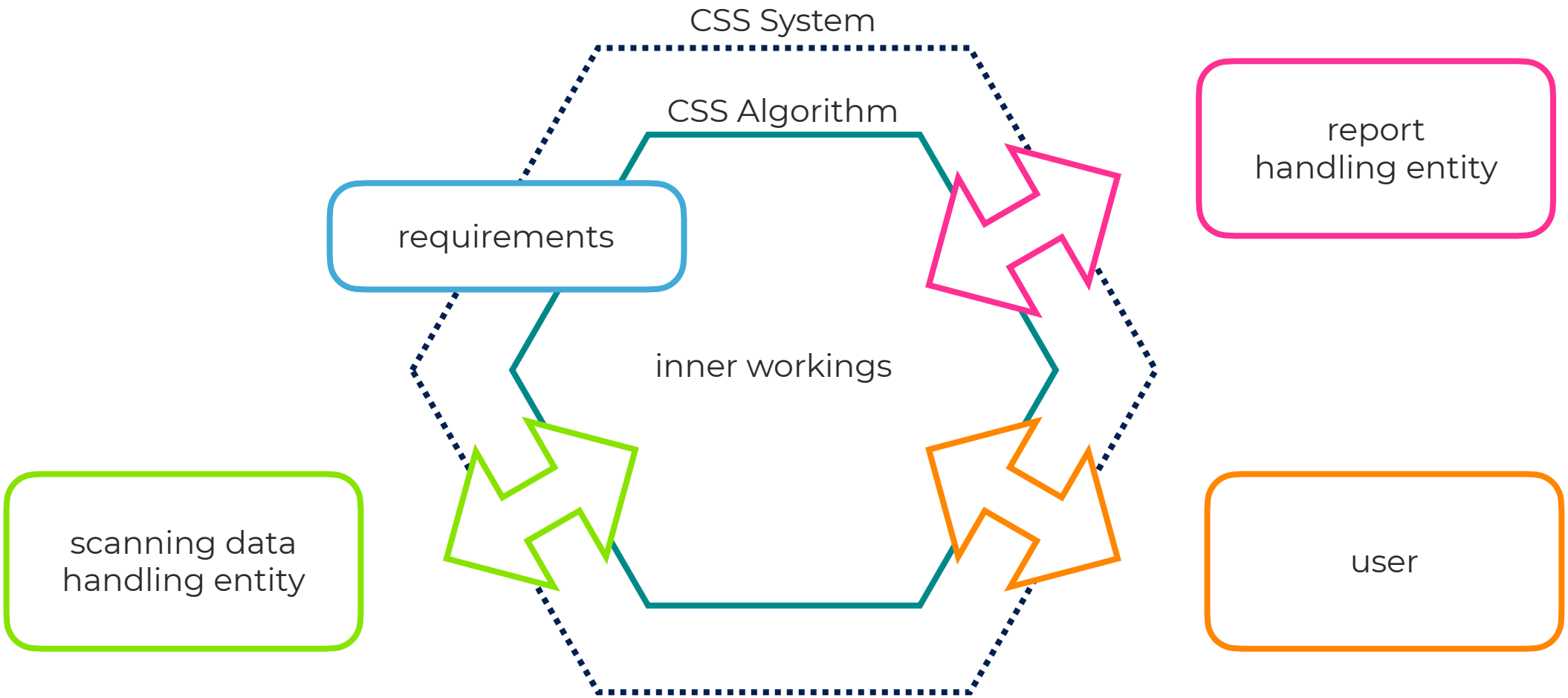
Results – Mental Models

decontextualized

Results – Mental Models

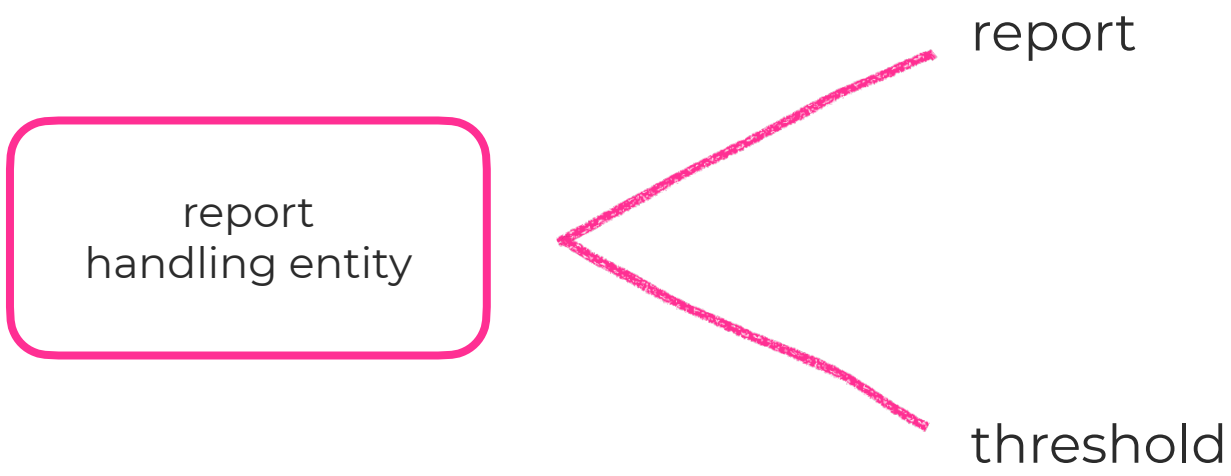
decontextualized

Results – Mental Models



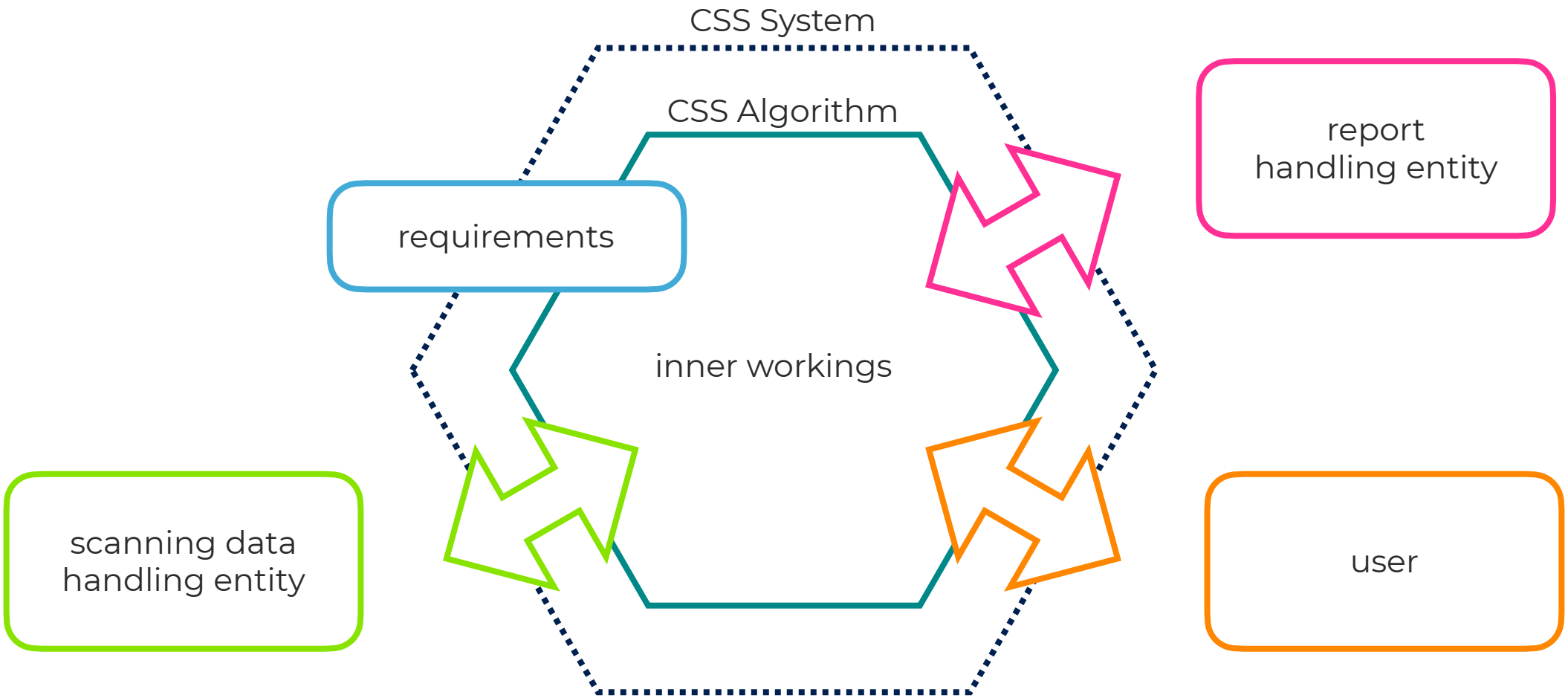
decontextualized

Results – Mental Models



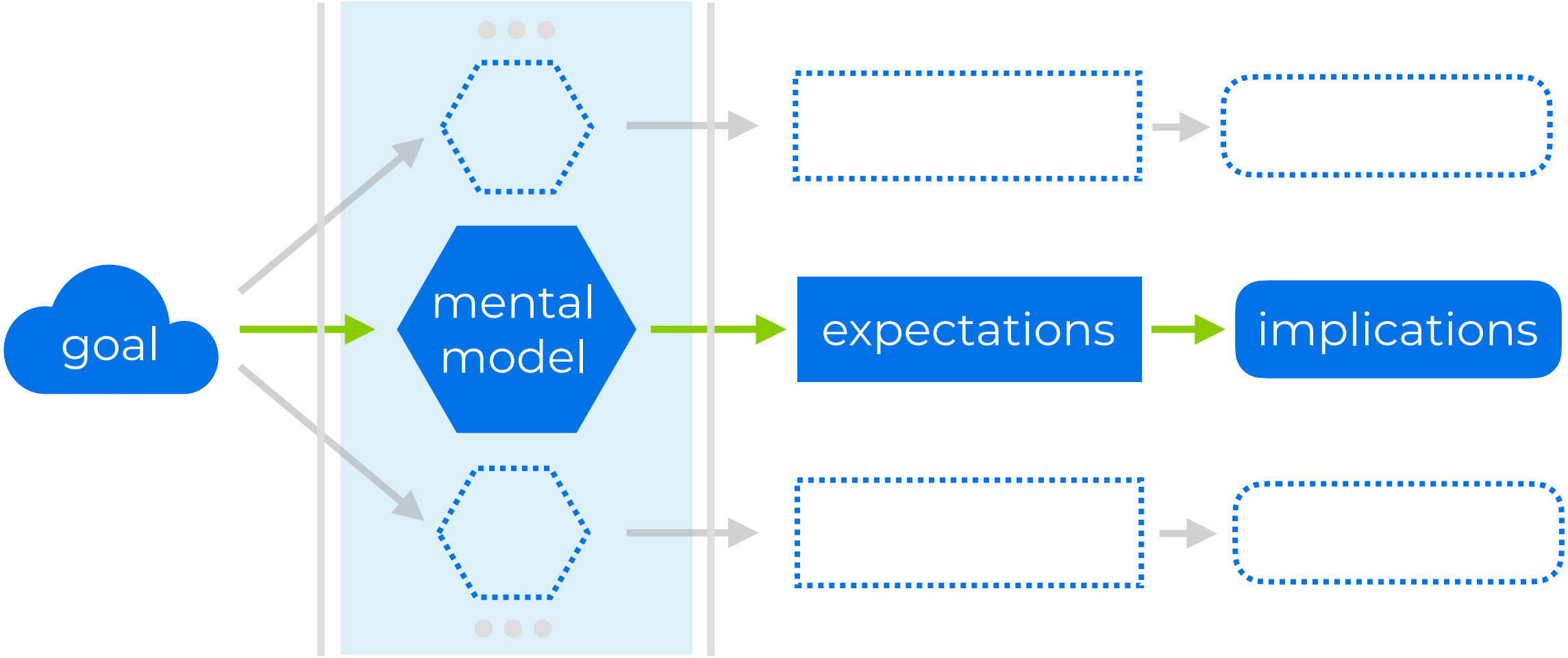
decontextualized

Results – Mental Models



decontextualized

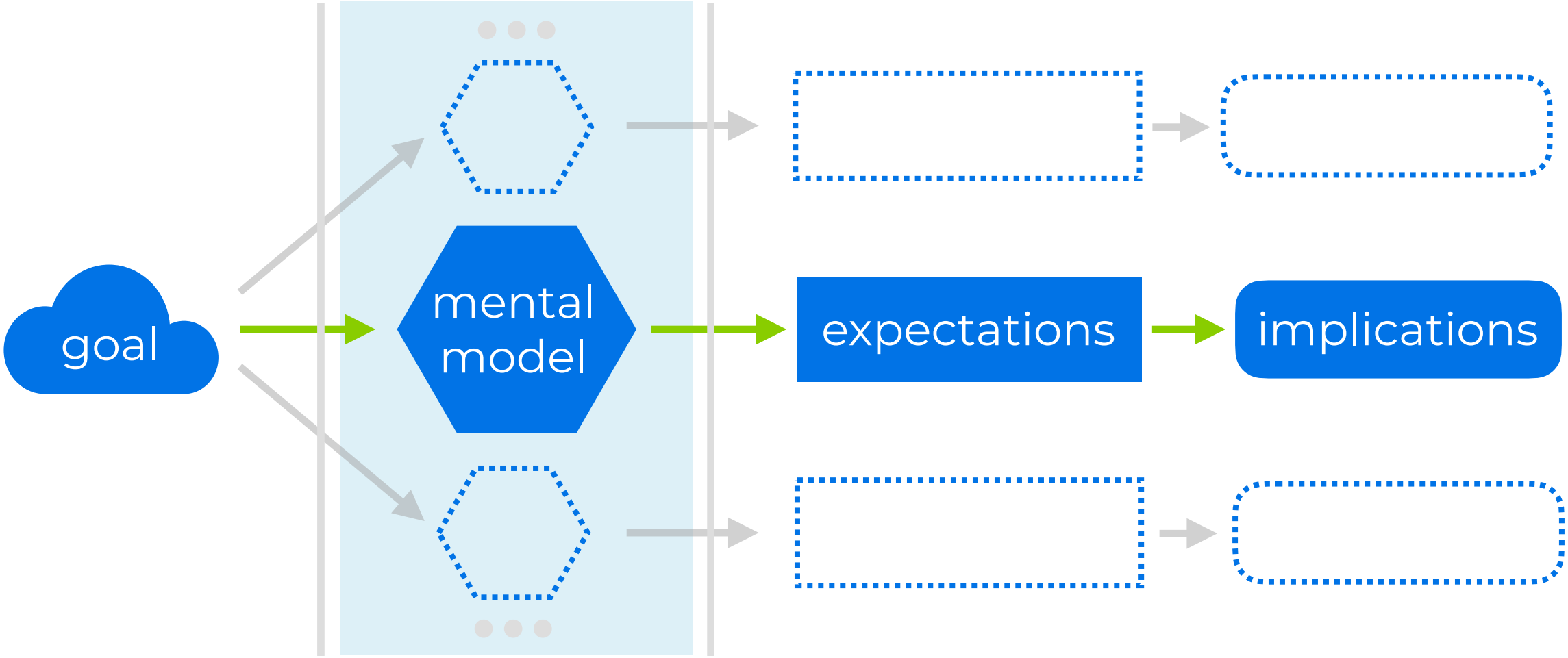
Results – Overview



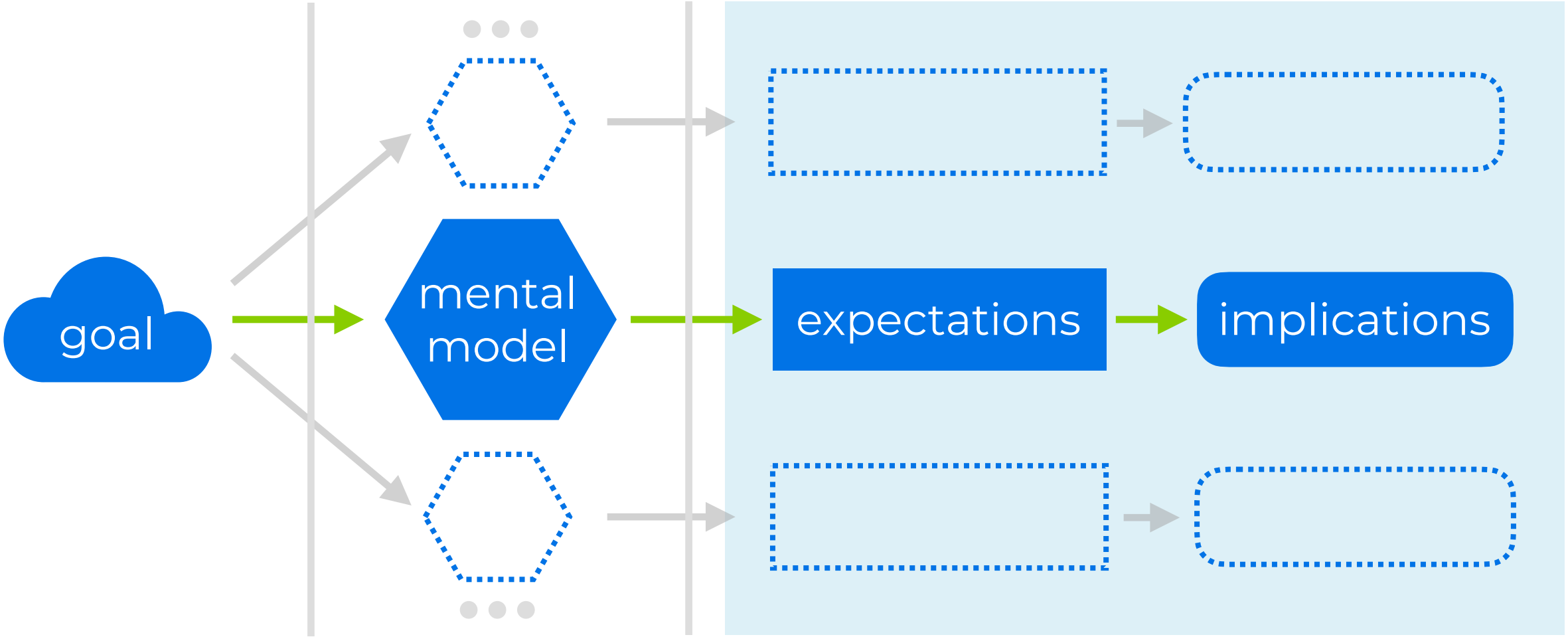
“In many ways, I think [CSS] is a bad solution that’s been created by a different problem.”

Participant X10

Results – Overview



Results – Overview



Results – Expectations

expectations

adoption,
reception

Results – Expectations

adoption,
reception

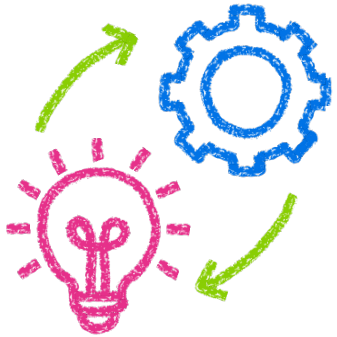
“[…] the lack of a good alternative then makes it attractive.”

Participant X07

Results – Expectations

expectations

adoption,
reception

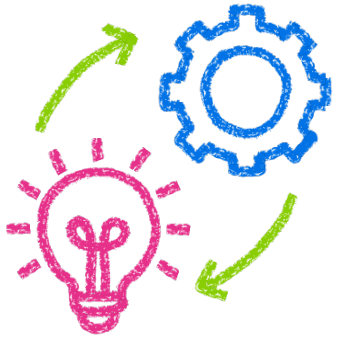


implementation

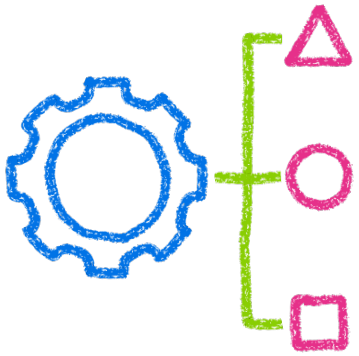
Results – Expectations

expectations

adoption, reception

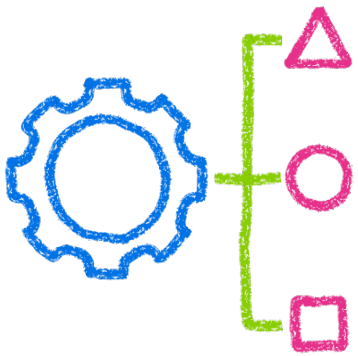


implementation



use or misuse

Results – Expectations



use or misuse

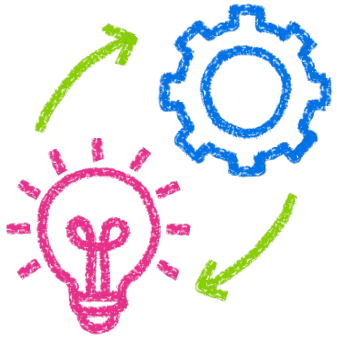
“It’s not linked to a particular domain at all, which is dangerous.”

Participant X03

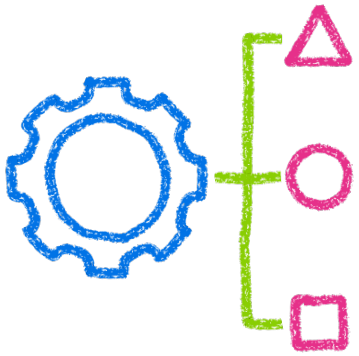
Results – Expectations

expectations

adoption, reception



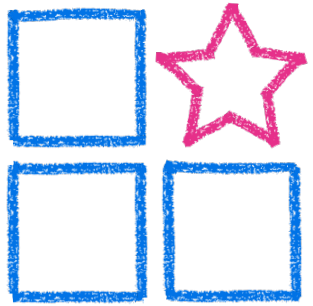
implementation



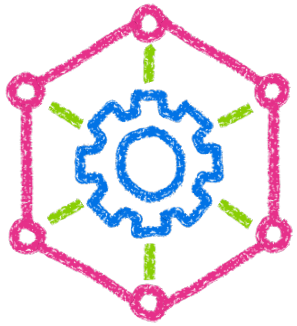
use or misuse

challenges

Results – expectations – Challenges



CSAM or
non-CSAM



integrity
of system

overwhelmed
legal system

Results – Implications

implications

personal

societal

threats

Results – implications – Stakeholders

children and teenagers

companies and service providers

DISCUSSION

Results – implications – Stakeholders

threat to children an teenagers

Results – implications – Stakeholders

user as the adversary

Results – implications – Stakeholders

Apple’s ‘Sensitive Content Warning’ Feature

Results – implications – Stakeholders

CSAM is a complex societal problem not easily solvable by technology

“I think an underlying problem of some of the Client-Side Scanning is this believe that technology will solve our societal problems. It won’t.

Participant X24

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

UNIVERSITÄT DES SAARLANDES

Divyanshu Bhardwaj★

Carolyn Guthoff★

Adrian Dabrowski

Sascha Fahl

Katharina Krombholz

★ both authors contributed equally

Mental Models, Expectations and Implications of Client-Side Scanning: An Interview Study with Experts

Divyanshu Bhardwaj‡
CISPA Helmholtz Center for Information Security, *and* Saarland University
Germany

Carolyn Guthoff‡
CISPA Helmholtz Center for Information Security, *and* Saarland University
Germany

Adrian Dabrowski
CISPA Helmholtz Center for Information Security
Germany

Sascha Fahl
CISPA Helmholtz Center for Information Security
Germany

Katharina Krombholz
CISPA Helmholtz Center for Information Security
Germany

What now?

- Current research: safety mechanisms for sexual risks in E2EE messengers with the goal of giving more agency to children, teenagers and adult users in navigating these sexual risks
- If you belong to one of the following groups and are interested in participating in research, please get in touch with me:
  - child protection
  - law enforcement
  - electronic service providers

Takeaways

- When building technology, built for users.
- Technology will not solve societal problems …
- … but with the right tools, we can help fight these problems.
  - Most forms of CSS are not the right tool to fight CSAM.

What impact would the introduction of a CSS System have on your company?

Carolyn Guthoff

carolyn.guthoff@cispa.de