ETH zürich

SECURITY
PRIVACY
SOCIETY

Better Together? - On the Resilience of Human-Technology Teams

Prof. Verena Zimmermann
Security, Privacy & Society
D-GESS

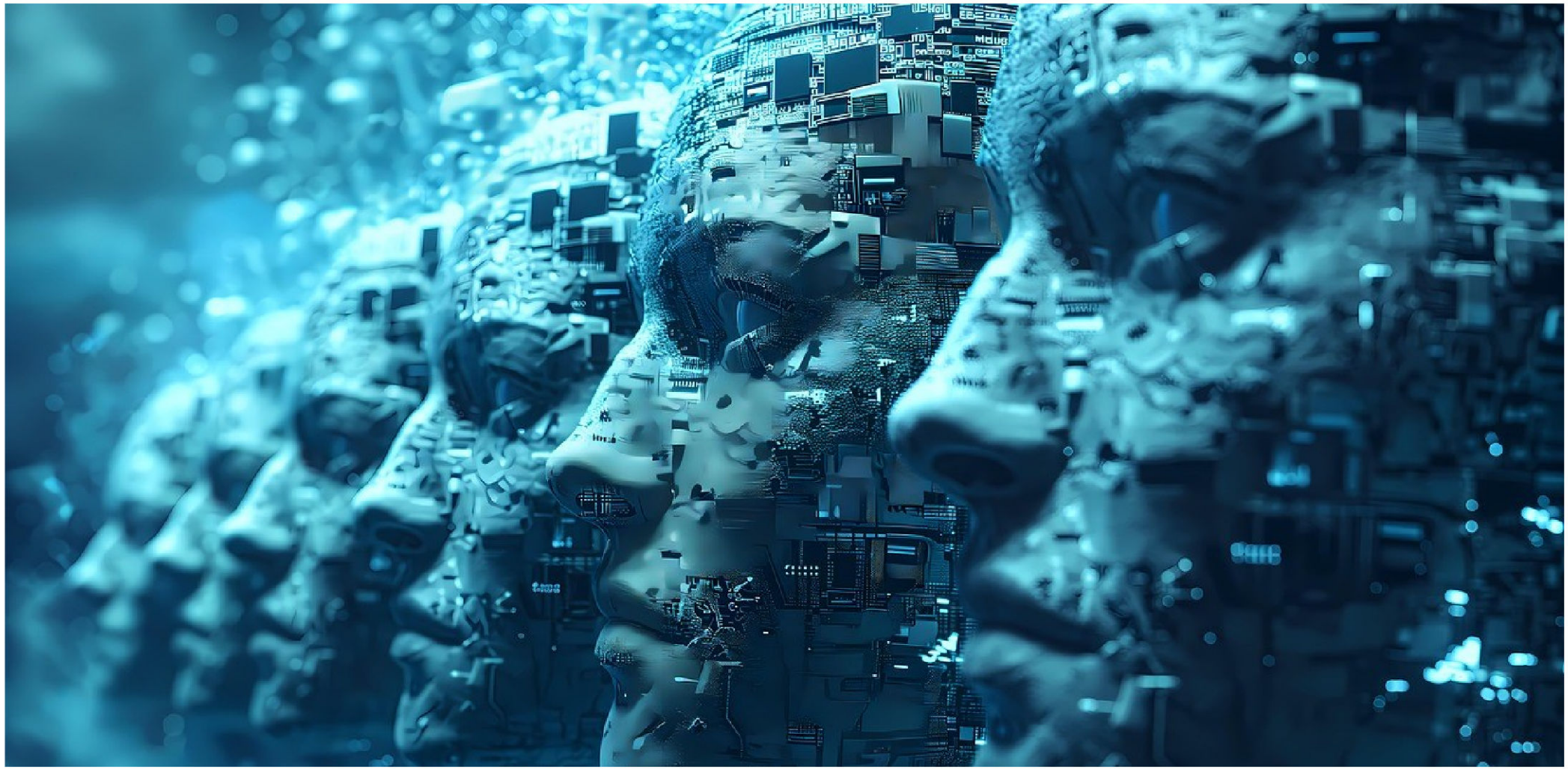
Human-centered Cybersecurity

SECURITY PRIVACY SOCIETY

- goes beyond making systems usable
- builds on psychological foundations
- enhances match between humans and technology
- fosters human contributions to security

Mission: from risk to resilience

Expert-AI Collaboration



Background

SECURITY PRIVACY SOCIETY

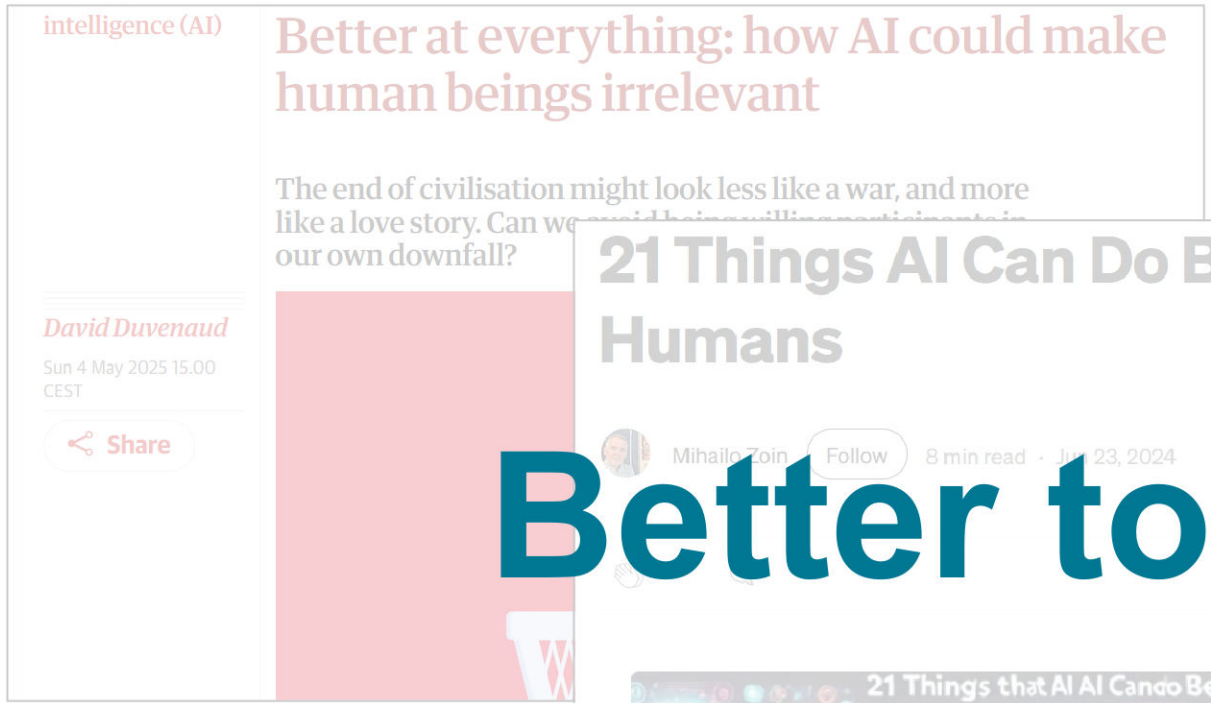
Cybersecurity is a complex field that constantly evolves and requires professionals with domain-specific knowledge and expertise

Cybersecurity professionals are a scarce resource and faced with high workloads

AI systems can handle data heavy landscapes quickly and efficiently

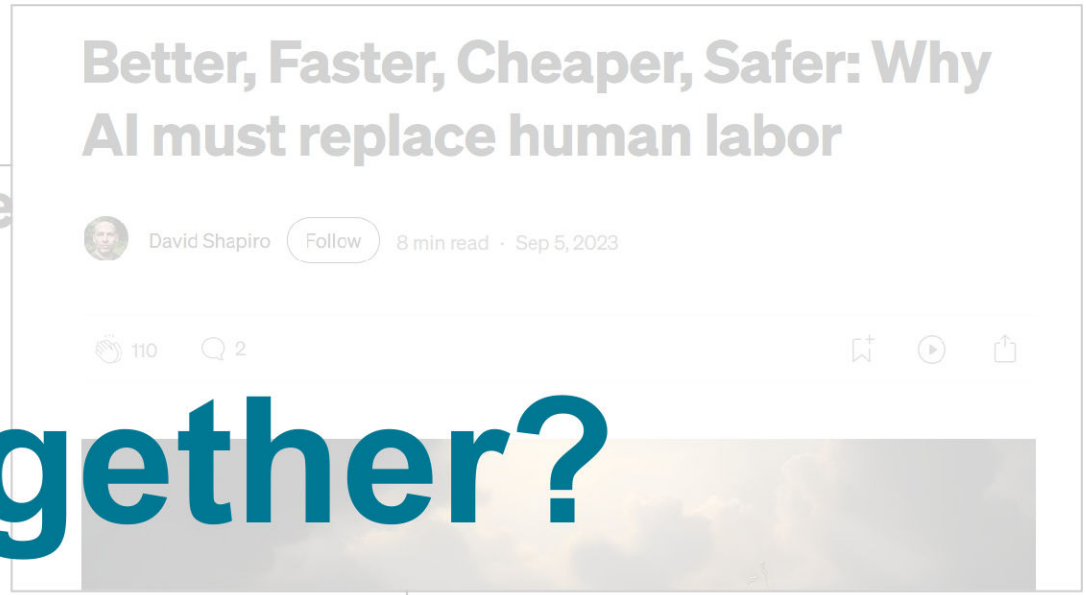
Experts and AI

SECURITY PRIVACY SOCIETY



https://www.theguardian.com/books/2025/may/04/the-big-idea-can-we-stop-ai-making-humans-obsolete

https://medium.com/@kombib/21-things-ai-can-do-better-than-humans-0a0cf389244f



https://medium.com/@dave-shap/better-faster-cheaper-safer-why-ai-must-replace-human-labor-20203020f5f7

Better together?

When combinations of humans and AI are useful (Vaccaro et al. 2024)

SECURITY PRIVACY SOCIETY

- Systematic review and meta-analysis of >100 experimental studies

Finding:
On average, human–AI combinations performed worse than the best of humans or AI alone

- **When humans outperformed AI alone** → performance gains in collaboration
- When AI outperformed humans alone → performance losses in collaboration

Augmenting Human Expertise with AI

SECURITY PRIVACY SOCIETY

Speed

Precision

Repetition

Data-driven decisions

Creativity

Emotional & social engagement

Handling uncertainty & adaptability

Experience & contextualization

But how can we design the collaboration in the high-stakes cybersecurity environment?

Together, professional and AI can

- perform complementary tasks
- use complementary skill sets,
- alleviate humans of workload,
- lead to better cybersecurity-related outcomes

Multi-Study Approach

SECURITY PRIVACY SOCIETY

Roch

Explore

Interview Study

Provide Transparency

Online Experiment

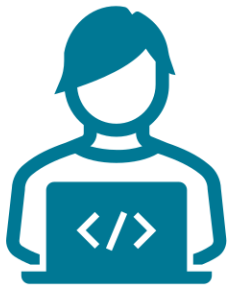
Adapt Autonomy

Online Experiment

Study 1: Explore
How to expert-AI collaboration for cybersecurity?

Interview study with 27 cybersecurity professionals

What are needs and requirements of practitioners for collaborating?



Professionals expressed

AI can improve their workflows

Relief from extensive and repetitive tasks

Help with more discretionary tasks

AI autonomy level depends on risk-benefit analysis

Study 1: Explore
How to expert-AI collaboration for cybersecurity?

SECURITY PRIVACY SOCIETY

AI Autonomy Levels (based on O'Neill et al., 2022)

Autonomous

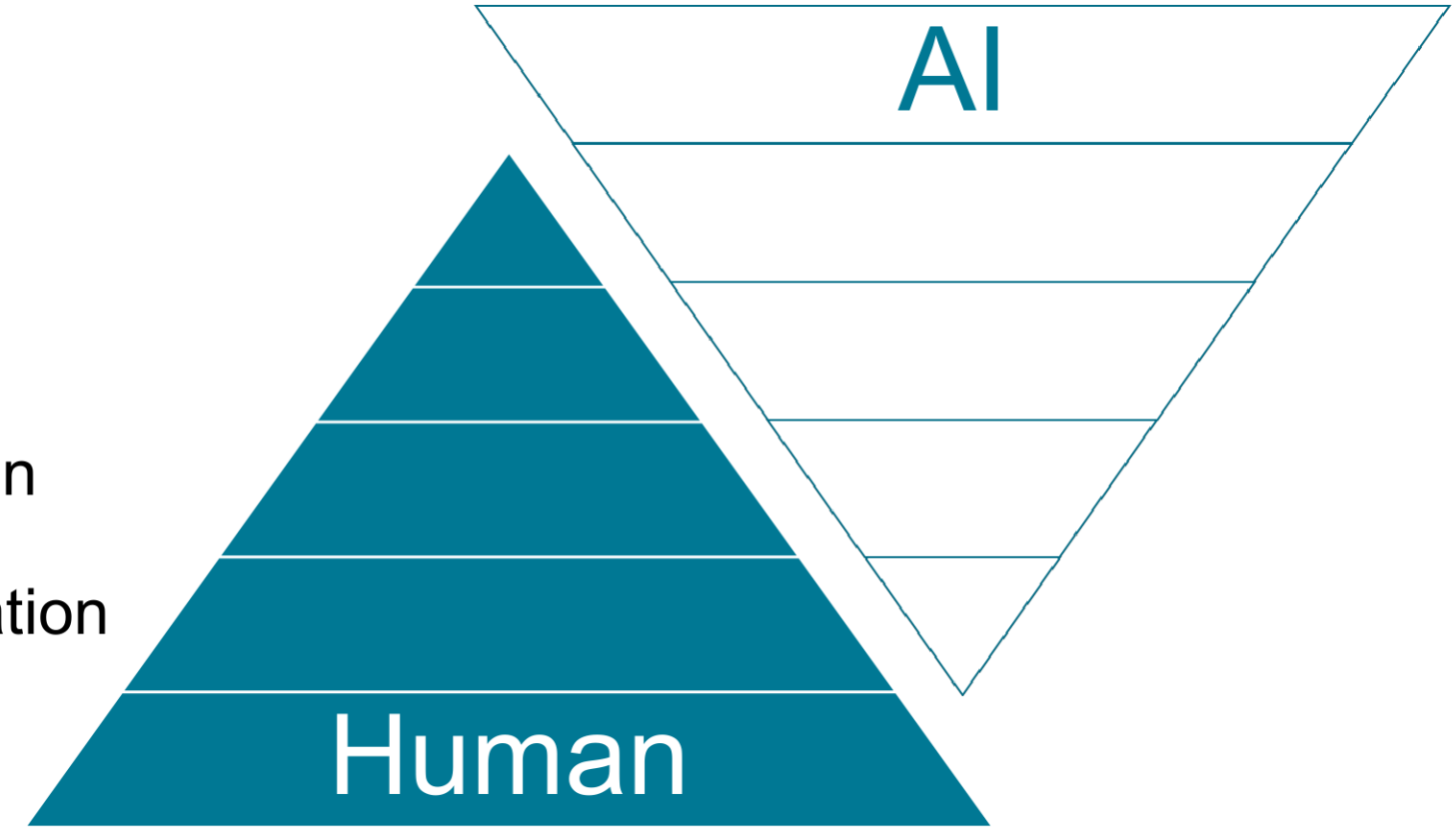
Human Veto

Human Approval

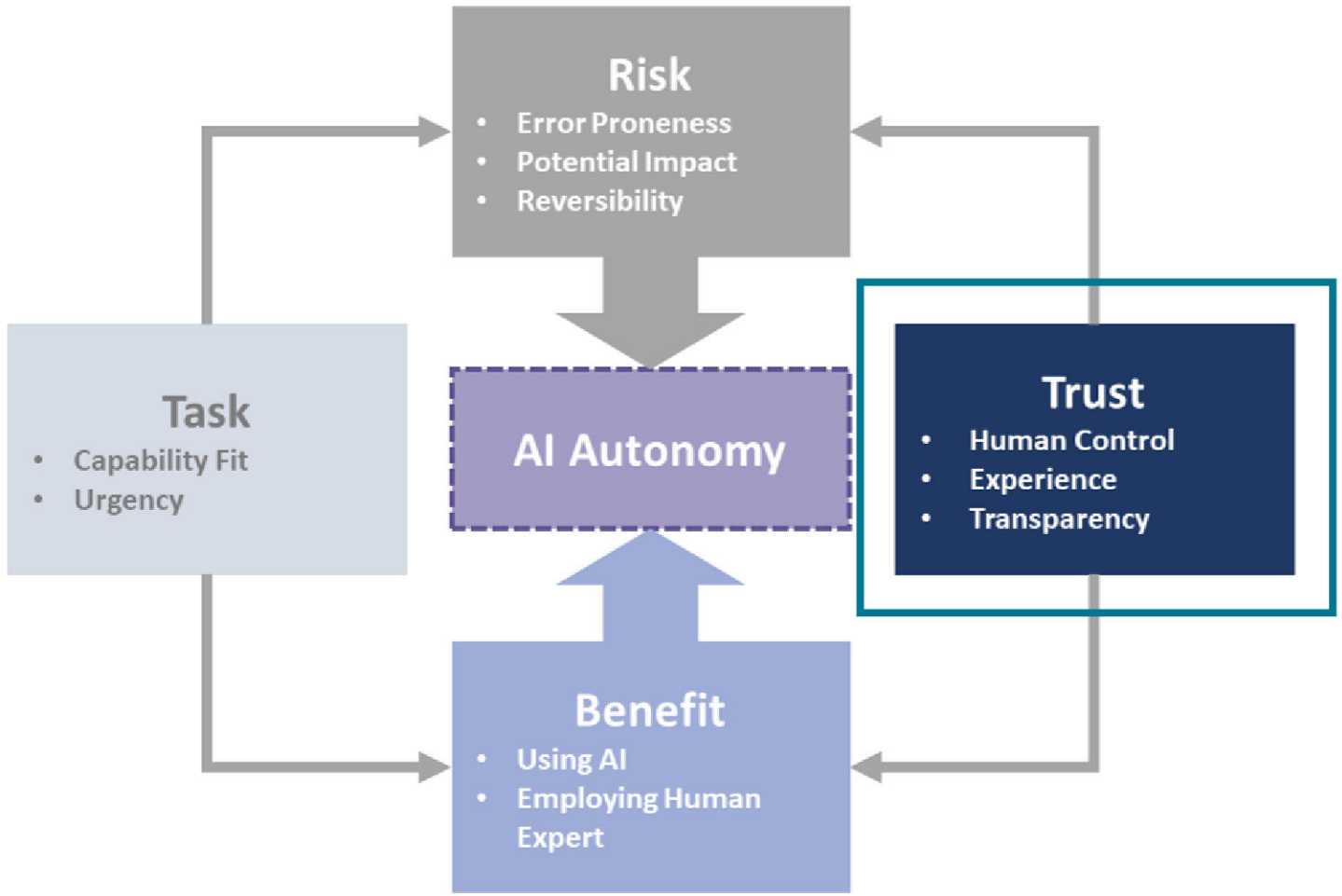
Information Analysis Automation

Information Acquisition Automation

Manual



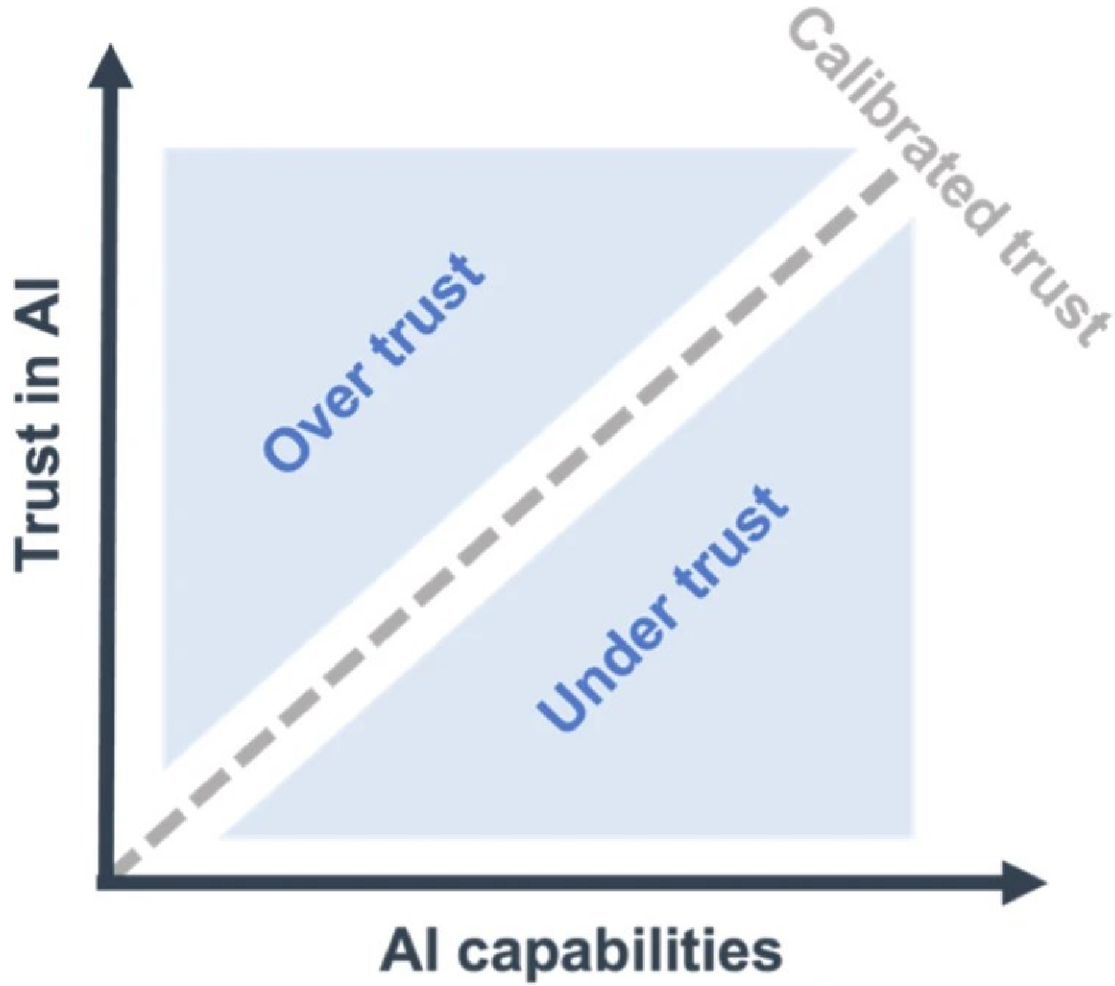
Study 1: Explore
How to expert-AI collaboration for cybersecurity?



Study 1: Explore
How to expert-AI collaboration for cybersecurity?

SECURITY PRIVACY SOCIETY

- Emphasis on **transparency** and need for a **trust relationship** with more autonomous AI in cybersecurity
  - Trust needs to be aligned with AI capabilities
  - Avoid humans overestimating AI systems' capabilities, leading to over-trust and -reliance → **misuse**
  - Avoid humans underestimating AI system capabilities, leading to under-trust and -reliance → **disuse**



Study 1: Explore
How to expert-AI collaboration for cybersecurity?

SECURITY
PRIVACY
SOCIETY

- Emphasis on **transparency** and need for a **trust relationship** with more autonomous AI in cybersecurity
  - Trust needs to be aligned with AI capabilities
  - Avoid humans overestimating AI systems' capabilities, leading to over-trust and -reliance → **misuse**
  - Avoid humans underestimating AI system capabilities, leading to under-trust and -reliance → **disuse**

Link to paper:

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY PRIVACY SOCIETY

- Selection of a relevant and critical cybersecurity task

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY
PRIVACY
SOCIETY

- Selection of a relevant and critical cybersecurity task
- Comparison of transparent vs. non-transparent design
  → **explainable AI (XAI)**

VS

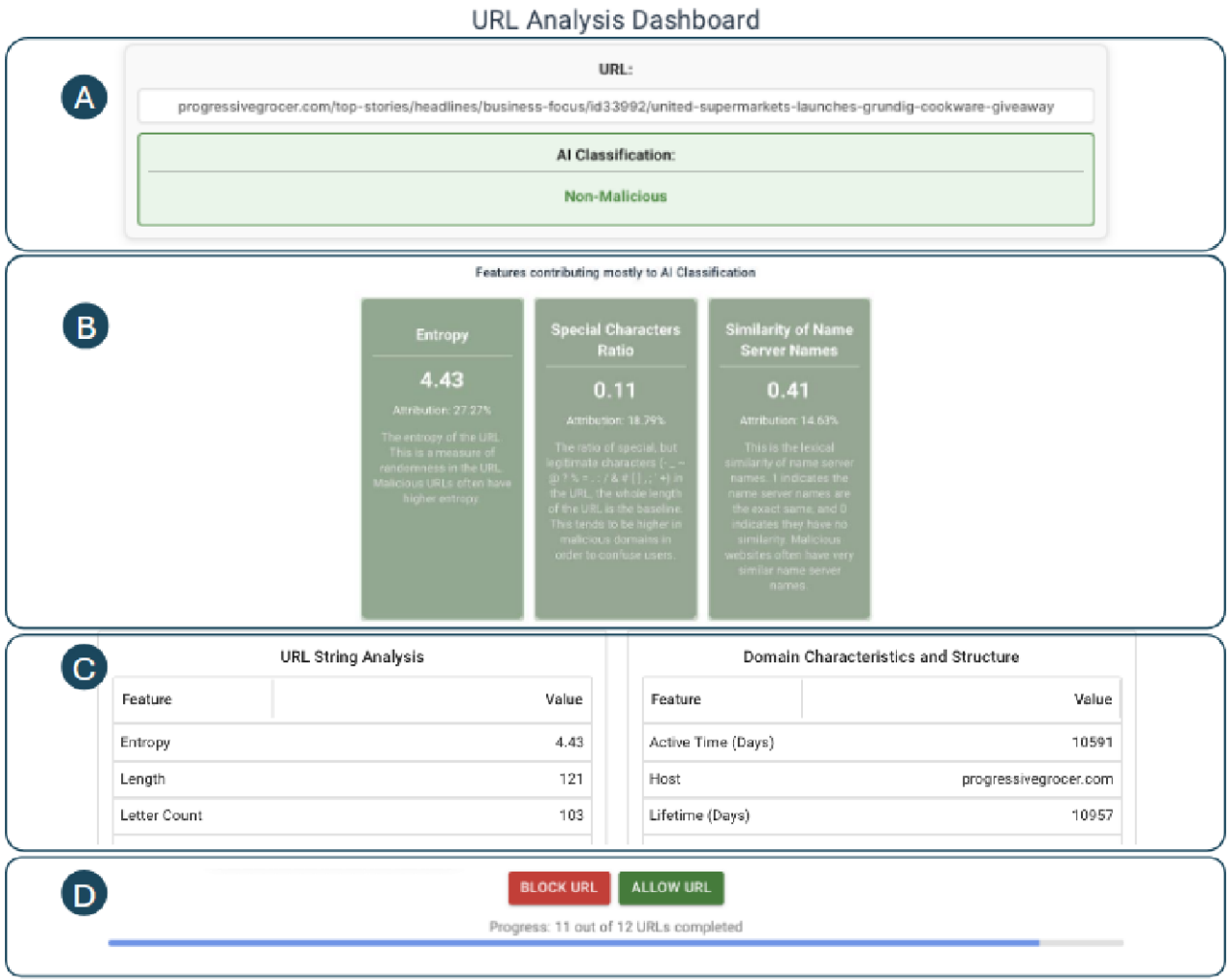
transparent

non-transparent

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY PRIVACY SOCIETY

- Selection of a relevant and critical cybersecurity task
- Comparison of transparent vs. non-transparent design
  → **explainable AI (XAI)**
- Providing explanations about the AI's decision-making in dashboard
- Online study with >100 cybersecurity professionals



Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY
PRIVACY
SOCIETY

Trust

Reliance

Performance
&Task Load

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY
PRIVACY
SOCIETY

Not able to foster appropriate trust

- Lower trust after interaction
- No improvements in recognising false AI classifications

Attitudinal trust and behavioural reliance do not align

No changes in performance or task load

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY PRIVACY SOCIETY

Not able to foster appropriate trust

- Lower trust after interaction
- No improvements in recognising false AI classifications

Attitudinal trust and behavioural reliance do not align

- Change in attitude but not behavior
- Overreliance

No changes in performance or task load

Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY PRIVACY SOCIETY

Not able to foster appropriate trust	• Lower trust after interaction • No improvements in recognising false AI classifications
Attitudinal trust and behavioural reliance do not align	• Change in attitude but not behavior • Overreliance
No changes in performance or task load	• Explanations did not increase task load but also not performance

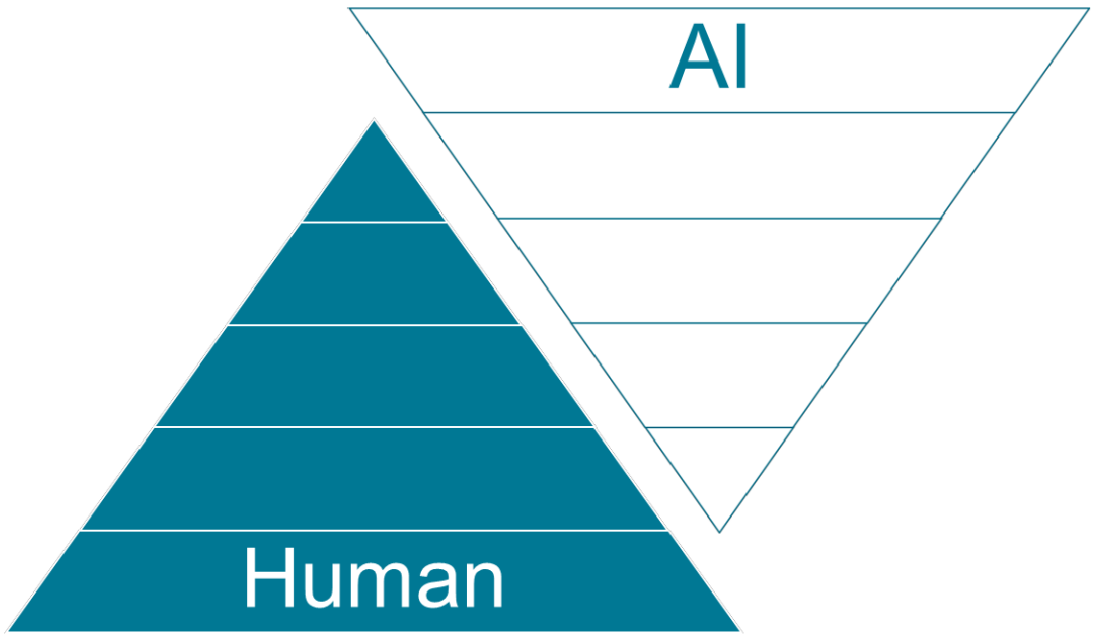
Study 2: Provide Transparency
Using transparency to enhance expert-AI trust and performance

SECURITY
PRIVACY
SOCIETY

Trust, reliance and performance do not depend on transparency per se but the complex interplay of the AI model and the expert's «model».

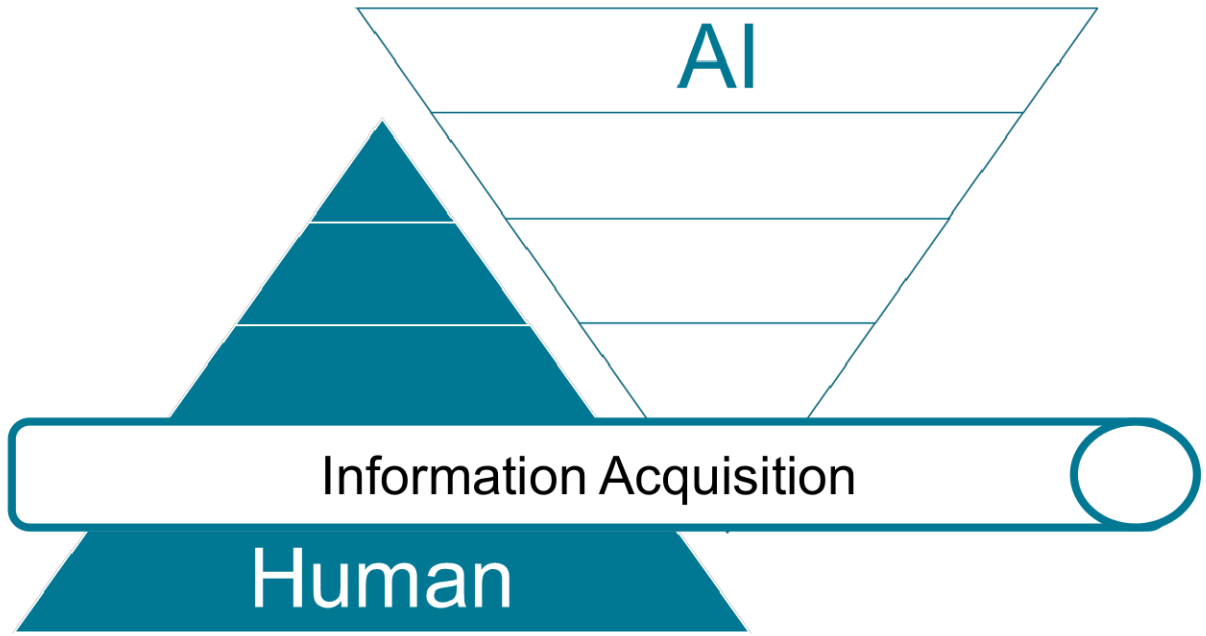
Study 3: Adapt Autonomy
Adapting AI autonomy for building appropriate trust (Ongoing)

- Make level of AI autonomy adaptable
- **Aims:**
  - provide control to practitioners over how collaboration should look
  - facilitate building appropriate trust
  - enhance human autonomy
  - foster situational awareness (ability to take back control if automation fails)



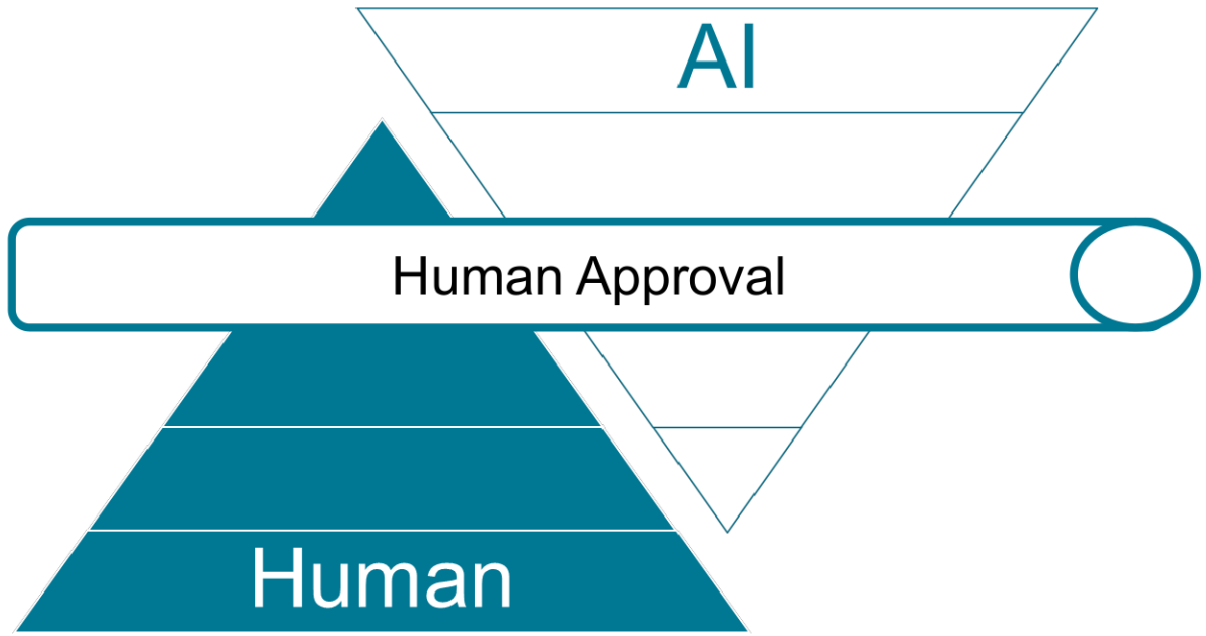
Study 3: Adapt Autonomy
Adapting AI autonomy for building appropriate trust (Ongoing)

- Make level of AI autonomy adaptable
- **Aims:**
  - provide control to practitioners over how collaboration should look
  - facilitate building appropriate trust
  - enhance human autonomy
  - foster situational awareness (ability to take back control if automation fails)



Study 3: Adapt Autonomy
Adapting AI autonomy for building appropriate trust (Ongoing)

- Make level of AI autonomy adaptable
- **Aims:**
  - provide control to practitioners over how collaboration should look
  - facilitate building appropriate trust
  - enhance human autonomy
  - foster situational awareness (ability to take back control if automation fails)



So… Better Together?

SECURITY PRIVACY SOCIETY

Yes…and no

- Identify **tasks in which humans excel** and outperfom AI alone

  → Augment that task with AI capabilities to **complement human**
- **Transparency** is relevant for but does not automatically increase trust

  → Consider **match between AI model and expert** «model»
- Trust builds over time and with «interaction experience»

  → Consider making AI autonomy levels adaptable to model human-human **trust building**
- Cybersecurity is a **high-stakes environment**

  → Consider relevance, suitability, and **criticality of task**

ETH zürich

SECURITY PRIVACY SOCIETY

See website:

Read paper:

Take part in an interview study:

References:

O'neill, T. et al. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, *64*(5), 904-938.

Roch, N. et al. (2024). Navigating autonomy: unveiling security experts' perspectives on augmented intelligence in cybersecurity. In *SOUPS 2024* (pp. 41-60).

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293-2303.