Clear, Actionable and Confidence-Inspiring Recommendations?

Comparative Study of AI-Generated and Human-Written PT Reports

Katarina Galanska · Maria Pibilota Murumaa · Vashek Matyas · Agata Kruzikova · Mike Just · Tomas Cerny

SOUPS 2026 · Hannover

MUNI FACULTY OF INFORMATICS

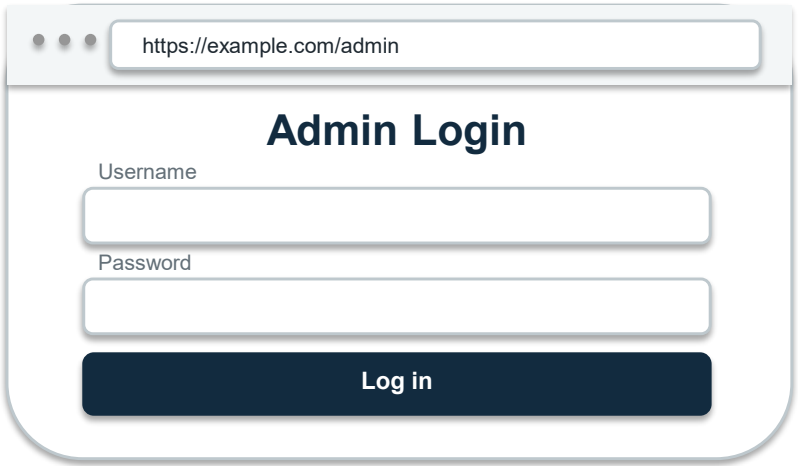
CROCS Centre for Research on Cryptography and Security

CHESS Cyber-security Excellence Hub in Estonia and South Moravia

Motivating scenario

Penetration Testing Finding

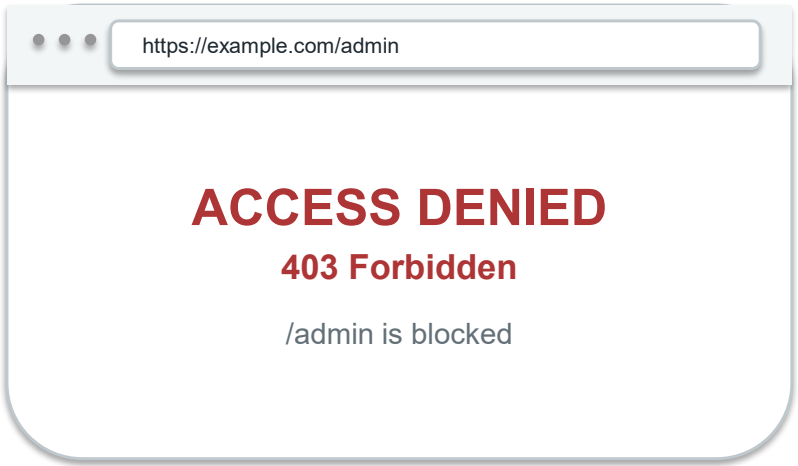
Administrative Interface Publicly Accessible



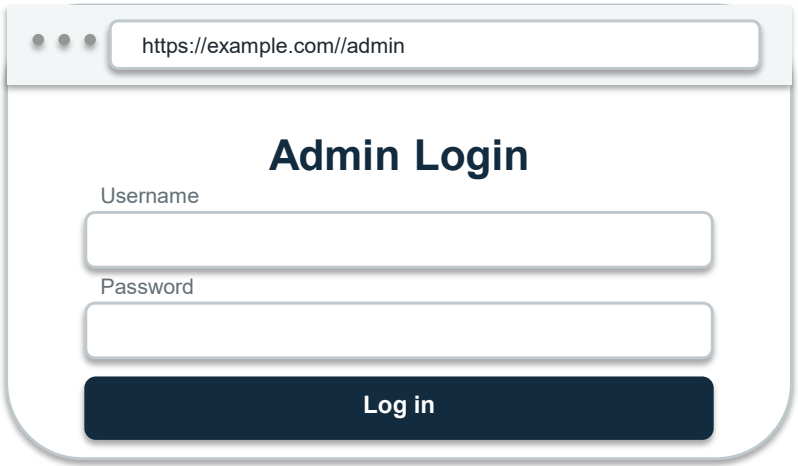
Motivating scenario

RECOMMENDATION **“Restrict access to the administrative interface.”**

1 Admin implements a path-specific restriction



2 Re-test uses an alternative path representation



Literal /admin request is blocked

Same interface is still reachable

The goal is clear, but what exactly should the recipient do?

STUDY DESIGN

How usable is the recommendation?

“Restrict access to the administrative interface.”

CLARITY

Is it easy to understand?

ACTIONABILITY

Is it clear what to do?

CONFIDENCE

Would I feel confident implementing it?

Human vs. AI · Level of detail · Responsibility Label

STUDY DESIGN

Five recommendation variants per vulnerability

Two vulnerability findings: information leak (I-type) and cross-site scripting (C-type)

10	20	30	40	50
Human-written	**AI · basic**	**AI · basic**	**AI · extended**	**AI · extended**
Pentester		+ role label		+ role label

DETAIL **basic → extended**

RESPONSIBILITY LABEL **no label → role label**

Participants and study design

598

participants

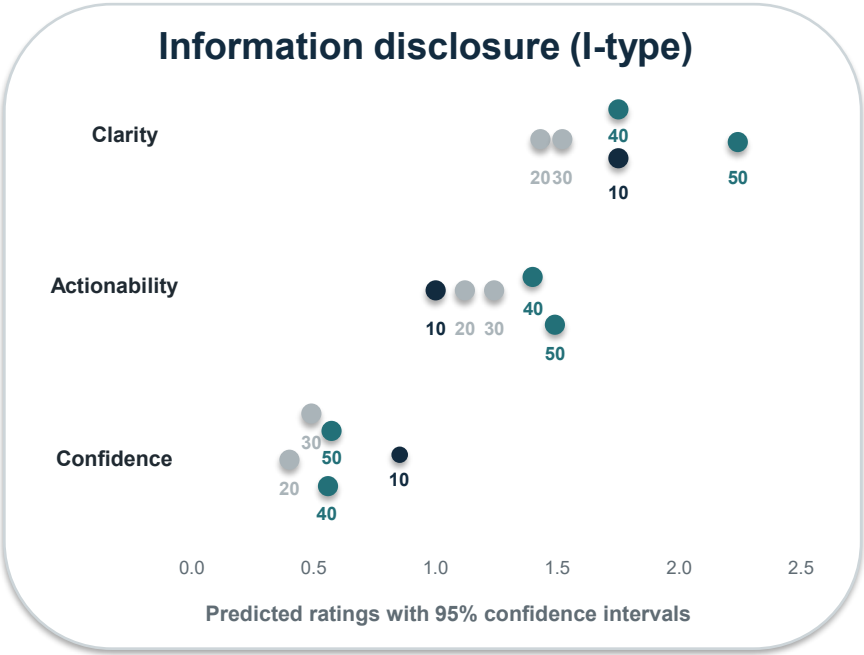
318	IT students
248	software-development professionals
32	security professionals

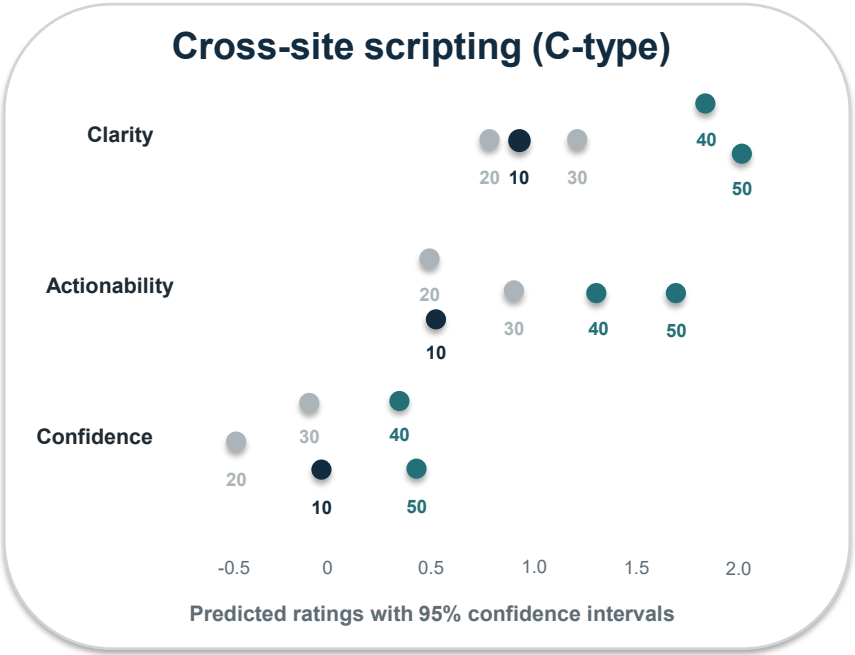


4-point scales for clarity, actionability, and confidence

Mixed-methods analysis: ordinal mixed models + thematic coding

Overall ratings: the pattern depended on the vulnerability





10 Human · 20 AI basic · 30 AI + role · **40 AI extended** · **50 AI ext. + role**

QUANTITATIVE RESULTS

The pattern depended on the vulnerability

Information Leak

Only two significant differences

CLARITY ↑

Basic → Extended + role

I20 → I50

Basic + role → Extended + role

I30 → I50

ACTIONABILITY & CONFIDENCE

no significant pairs

Cross-Site Scripting

13 significant pairwise differences overall

Basic AI → Extended AI

C20 → C40

Clarity ↑

Actionability ↑

Confidence ↑

Basic AI → Extended + role

C20 → C50

Clarity ↑

Actionability ↑

Confidence ↑

Largest observed difference: C20 vs. C50.

QUALITATIVE RESULTS

Why did participants prefer a recommendation?

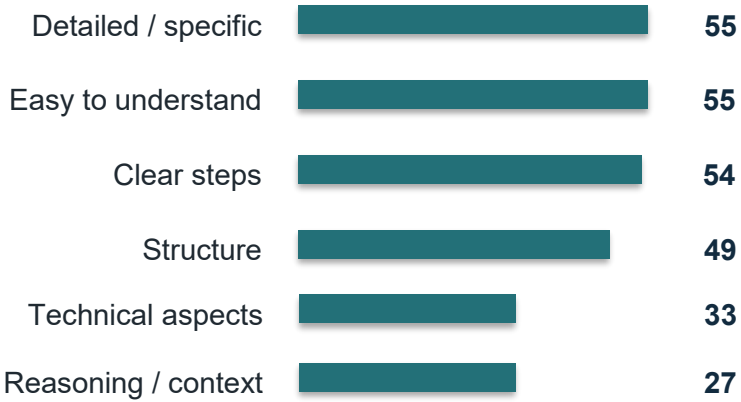
598
PARTICIPANTS
overall mixed-methods study

506
EXPLANATIONS
291 clarity + 215 actionability

1 code
FREQUENCY
how often a reason appeared

2 codes
CO-OCCURRENCE
how often two reasons appeared together

CLARITY · TOP 10 code frequencies



ACTIONABILITY · TOP 10 code frequencies



Which reasons appeared together?

CLARITY · 291 explanations

Structure — 13× — Clear steps

strongest reported pairs

Structure — 13× — Easy to understand

Clarity links structuring with clear steps and understanding.

ACTIONABILITY · 215 explanations

Technical aspects — 16× — Feels secure / risk mitigated

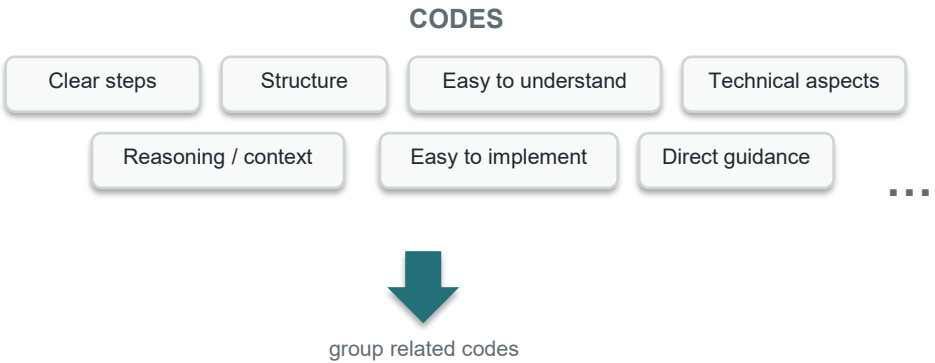
strongest reported pair

Clear steps — 14× — Easy to implement

Clear steps — 14× — Direct / straightforward

Technical aspects — 13× — Practical

What broader patterns emerged?

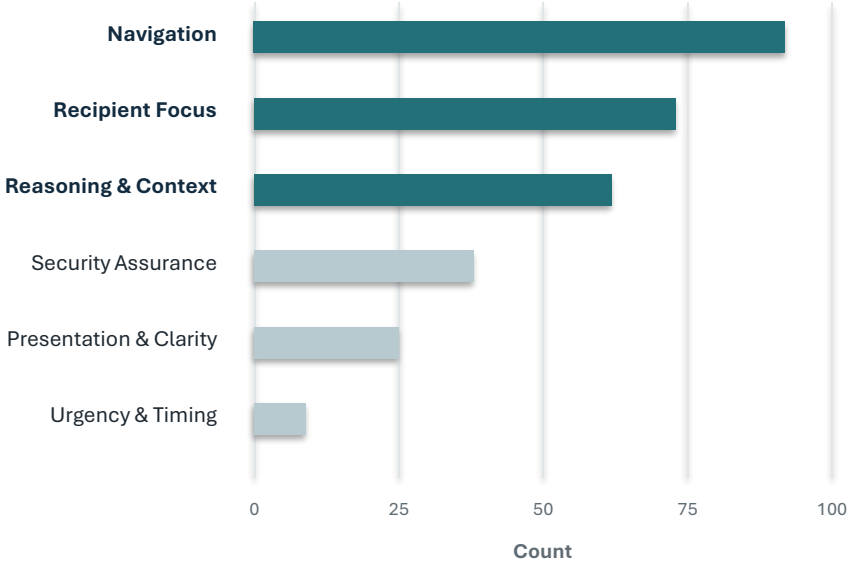


QUALITATIVE RESULTS

The same top themes matter, but in a different order

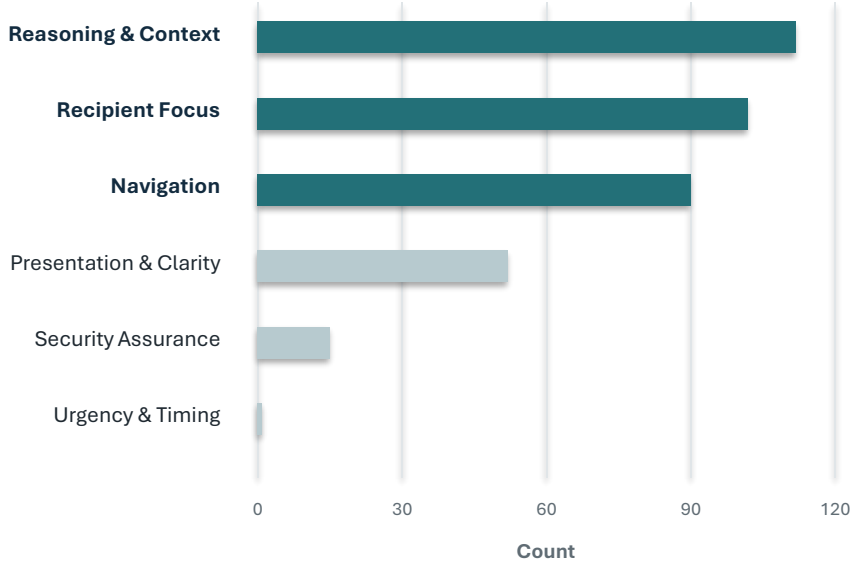
CLARITY

291 qualitative explanations



ACTIONABILITY

215 qualitative explanations



Navigation · Recipient Focus · Reasoning & Context

Clarity and actionability emphasize them differently.

INTERPRETATION

What mattered?

H ↔ AI

AUTHORSHIP

No simple human-vs-AI story

AI-generated recommendations often received higher ratings, but not uniformly across vulnerabilities and for all aspects.

+

LEVEL OF DETAIL

Extended guidance can help

High-detail variants were generally rated higher. Participants valued clear steps, structure, technical grounding, and reasoning or context

@

ROLE TAG

Higher Predicted Ratings

Versions with explicit responsibility information did consistently have higher predicted rating for clarity, actionability, or confidence.

CONCLUSION

Anatomy of usable recommendation

RECIPIENT FOCUS

STRUCTURE
CLEAR STEPS

BACKEND DEVELOPER

Restrict access to the administrative interface.

1. Enforce authorization for administrative routes.
2. Apply the restriction consistently before routing.
3. Verify that alternative path representations do not bypass the control.

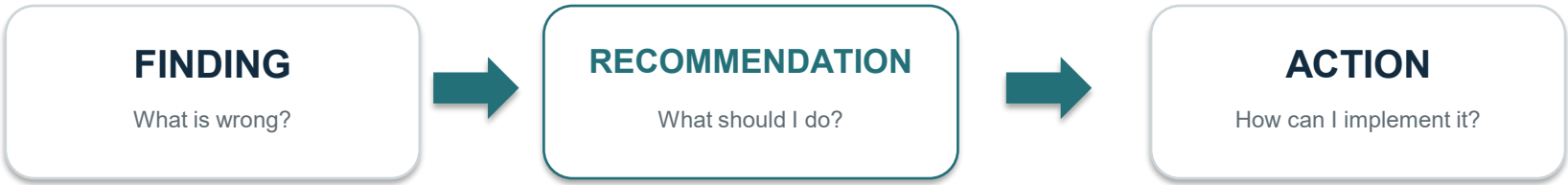
Why: prevent unauthorized users from reaching administrative functionality.

DETAIL

REASONING
DIRECTNESS

From Finding to Action

Make the recommendation easy to translate into practice



Thank you!