AI-SS European Dependable Computing Conference (EDCC) IEEE 2026

University of Kent

Resilience to Disinformation

(among cyber security and information warfare professionals)

Why Some Experts May Get It Wrong

Allison Wylde[1], James Fabio Petani[2], Hema Issa[3] and Fortuna Casoria[2]

1.Institute of Cyber Security for Society, University of Kent; email aw102@kent.ac.uk

2.Burgundy School of Business, CEREN EA 7477, Université Bourgogne Europe Dijon, France

3.ESSCA School of Management, Boulogne-Billancourt, CEDEX,Paris, France

Background and Problem Statement

The Stakes

Disinformation is the **#2 global threat in 2026** (World Economic Forum (WEF) Global Risks Report, 2026).
State-sponsored campaigns increasingly **target professional communities,** not just the public. Generative AI has dramatically increased volume and sophistication.

The Research Gap

- Most research to date has examined general public susceptibility, not expert practitioners
- Experts are assumed resilient; though the assumption is not thoroughly tested
- Expert practitioners may face unique vulnerabilities; overconfidence, overwork...

RQs: How often are professionals deceived?
What countermeasures do they use? How confident are they in detection?

Literature Goal

Construct and conceptual clarity essential: concepts, definitions and meaning vary across different disciplines (cf. Suddaby, 2010).

Aim to create foundations and building blocks based on concrete terms to:
allow analysis; create technical solutions (policy enforcement engines); enable theory building;
and, inform policy-making.

- 1. Disinformation, 2. Trust and 3. Decision-making

1. Disinformation

US “deliberately created to mislead, harm, or manipulate a person, social group or country” (CISA, nd.).

EU, “false or misleading content that is spread with an intention to deceive or secure economic or political gain, and which may cause public harm” (EU, 2018).

Literature

2. Trust

Management, conflict management and organization studies literature.

Definition; confident positive expectations of another's behaviour (Lewicki et al., 1998) and a willingness to be vulnerable irrespective of control or monitoring (Mayer et al., 1995).

Trust is subjective; multi-dimensional; multi-level; and, multi-referent.

3. Decision-making

In this paper, decision-making (DM) is conceptualised as whether an individual trusts or distrusts a particular individual, organization or factor.

Elements of expert DM draw on tacit knowledge (Nonaka & Takeuchi, 1995).

Who We Surveyed, Methodology

Cyber Security 38.8%

AI Research 34.7%

Info Warfare 19.4%

Policy & Gov. 26.5%

Comms & Media 23.5%

N=98, many respondents posses and operate across multiple expertise domains

Research approach

Structured questionnaire
Purposive sampling; professional networks and conferences
Single-item ordinal scales
Cross-sectional design

University ethics approval obtained; voluntary participation
Exploratory inferential analysis added post-review
Part of a larger study, interviews to follow

Meet the Respondents (N=98); Primary Domain of Expertise

Cyber Security 38.8%

AI Research 34.7%

Info Warfare 19.4%

Policy & Gov. 26.5%

Security Comms & Media 23.5%

Many respondents posses and operate across multiple expertise domains

Experience

<1 yr 2%
1–3 yr 16.3%
4–7 yr 16.3%
8–15 yr 31.6%
>15 yr 31.6%
(2.2% no response (nr))

Gender

67.3% male
21.4% female
4.1% non-binary/
prefer not to say
(7.2% nr)

Age

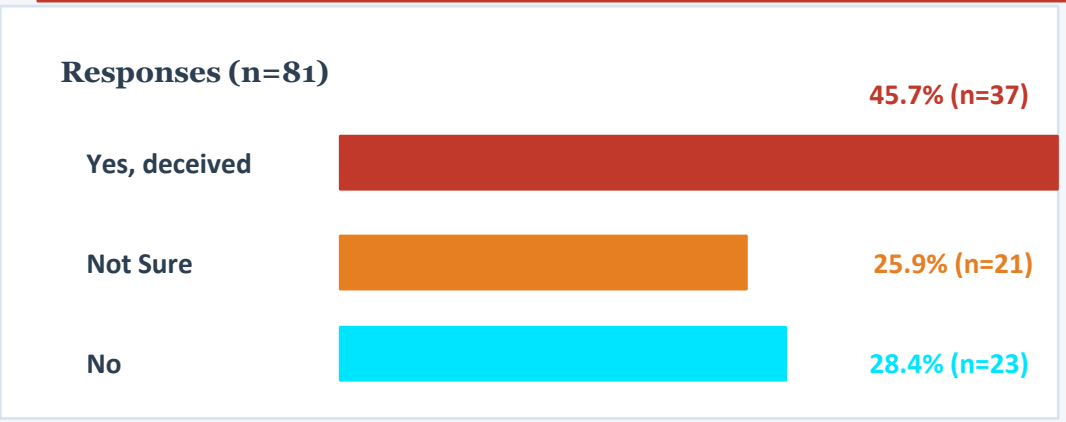
18–24 1%
25–34 22.4%
35–44 33.7%
45–54 24.5%
55–64 11.2%
65+ 4.1%
(3.1% nr)

Key Finding #1, Victimisation by Disinformation

45.7%

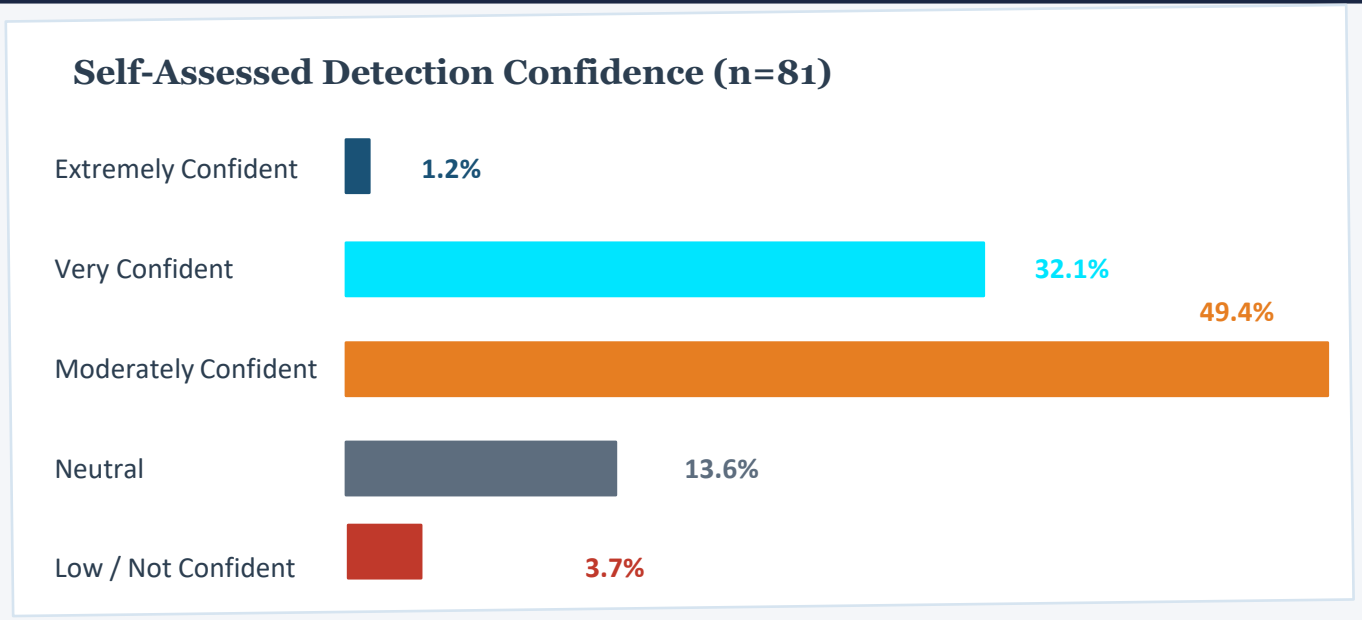
of valid respondents (n=81) acknowledged being deceived into sharing false information

a sobering finding for a domain-expert population



Key insight: 71.6% were either deceived or unsure; suggests actual vulnerability exceeds consciously recognised deception. That 25.9% 'replied ‘unsure' response suggests evidence of undetected manipulation.

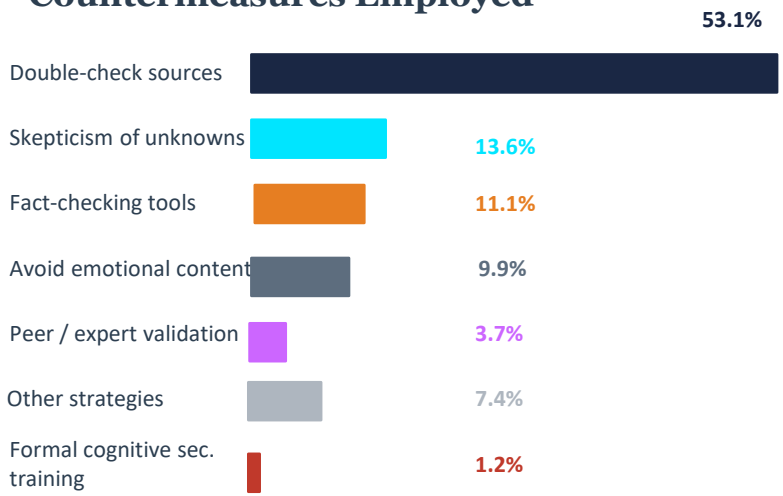
Key Finding #2, Detection Confidence



Calibration gap: Experience correlates with confidence (ρ=0.44, p<.001); but also with victimisation. Senior professionals are **more confident yet still acknowledge deception**, consistent with the 'illusion of explanatory depth'. Moderate confidence (49.4%) may reflect appropriate humility among those who have already been ‘burned’.

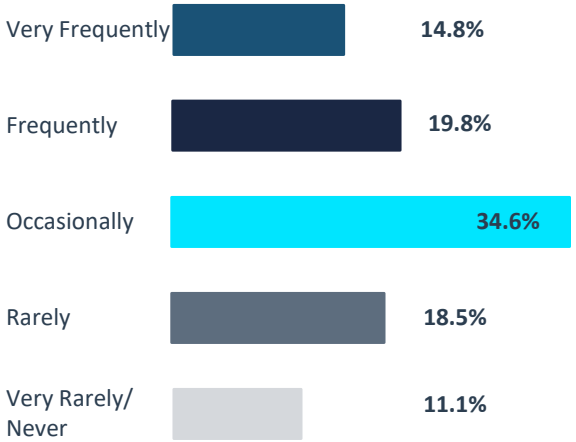
Key Finding #3, Countermeasures and Encounter Frequency

Countermeasures Employed



⚠ **Only 1.2% completed formal cognitive security training; the most critical gap identified**

Encounter Frequency



69.2% encounter disinformation regularly

Theoretical Contribution: Four-Factor Vulnerability Model

01

Information Overload

High-volume environments exceed cognitive verification capacity, increases the probability that false content bypasses scrutiny (Simon, 1971).

02

Trust Network Exploitation

Disinformation entering professional networks via authoritative sources gains elevated baseline credibility (Benkler et al., 2018).

03

Adversarial Sophistication

State-sponsored and AI-augmented operations engineered to evade expert heuristics. LLMs enable mass-personalised attacks (Goldstein et al., 2024).

04

Confirmation Bias

Schema-consistent disinformation accepted preferentially, even among analytically sophisticated recipients (Nickerson, 1998).

The 4-factors interact and compound: trust exploitation is most effective when content is also schema-consistent; overload reduces cognitive resources for lateral verification.

Statistical Analysis

Exploratory analyses on reconstructed ordinal data, hypothesis-generating only (α=0.05, two-tailed)

Test	Variables	Statistic	p-value	Interpretation
Spearman E1	Confidence × Experience	ρ = 0.44	< .001 ***	More experience → higher confidence
Spearman E2	Victimisation × Experience	ρ = −0.37	.001 **	More junior → higher acknowledged deception
Chi-sq E3	Victimisation × Exp. Group	$\chi^2(2) = 5.68$	.058 (ns†)	Trend only: 73% junior vs 39% senior deceived
Chi-sq E4	Confidence × Exp. Group	$\chi^2(1) = 4.51$	.034 *	Senior significantly more confident

Experience raises confidence; but senior professionals ALSO acknowledge victimisation, suggesting confidence, this accuracy gap persists across career stages ('illusion of explanatory depth').

Higher acknowledged deception among juniors (ρ=−0.37) may reflect greater self-awareness rather than greater susceptibility; senior deception may simply be underecognised.

Chi-square trend (E3, p=.058) is underpowered (n=15 junior). Fisher's exact test recommended, note caution in interpretation.

Limitations, Critical Appraisal

Self-Report Bias

High

Victimisation rates are self-assessed; actual deception likely exceeds acknowledged deception. The 25.9% 'unsure' response is evidence of undetected manipulation.

Self-Selection Bias

Moderate

Purposive sampling through disinformation-aware networks likely over-represents heightened-awareness professionals, inflating encounter rates and countermeasure awareness.

Cross-Sectional Design

Moderate

Causal inference is not possible. Longitudinal or experimental designs needed to establish whether training, experience, or context drive observed

Single-Item Scales

High

All constructs measured with single items. No established reliability or validity. Multi-item validated instruments required.

No Credential Verification

Moderate

Professional eligibility relied on self-identification. No independent verification of domain expertise conducted.

Inferential Constraints

Moderate

Tests applied to reconstructed aggregated data, not raw microdata. Junior subsample (n=15) underpowered Larger sample recommended.

These limitations bound, but do not invalidate the study's contribution; calls for replication with larger, verified samples and validated instruments.

Implications and Conclusion

1

Mandatory Cognitive Security Training

1.2% training uptake is a critical failure. Inoculation-based and lateral reading must become standard CPD across all expert domains.

2

Institutional Verification Protocols

Individual vigilance insufficient. Organisations must implement systematic, workflow-embedded verification to reduce cognitive load.

3

AI-Assisted Detection Tools

Human analysis cannot match the volume of AI-generated disinformation. Automated detection must complement, not replace, expert judgement.

4

Cultural Openness About Vulnerability

Reduce stigma around admitting deception promotes defensive information-sharing and collective organisational resilience.

Conclusion: 45.7% of cyber security and IW professionals acknowledged deception; 69.2% encounter disinformation regularly yet rely on informal defences. The four-factor model explains why expertise alone is insufficient protection. Future research requires larger verified samples, validated multi-item instruments, and longitudinal designs tracking training effects.

Future Research Agenda

- Larger verified samples to enable domain-specific subgroup analyses and sufficient inferential power
- Multi-item validated instruments for each construct (victimisation, confidence, countermeasures)
- Longitudinal designs tracking effects of cognitive security training interventions over time
- Experimental / behavioural measures to complement self-report data
- Investigation of AI-augmented disinformation specifically engineered to exploit expert heuristics

The accelerating integration of generative AI in disinformation operations makes this research agenda urgent and consequential

Thank you

Dr Allison Wylde FHEA FRGS IEEE
Visiting Academic
University of Kent, School of Computing
Institute of Cyber Security for Society
Canterbury, Kent, UK

aw102@kent.ac.uk
https://research.kent.ac.uk/cyber/person/allison-wylde/
https://orcid.org/0000-0002-1121-1685