From Prompt to Propaganda: Mapping the Emerging Disinformation Lifecycle and Research Ecosystem

An Interdisciplinary Workshop-Grounded Cartography

Marc-Oliver Pahl[1]

Abstract: Generative AI changes not only the production of misleading content but the economics and organisation of influence operations. This paper synthesises an interdisciplinary workshop bringing together computer science, cybersecurity, journalism, social science, governance, national counter-disinformation ecosystems and professional communities. Using proceedings papers, reconstructed expert interventions and contextual sources, we map capabilities onto a lifecycle from planning and targeting to generation, dissemination, amplification and impact. Three refinements emerge: analysis should expand from artefacts to operations; information and audience-data flows interact; and influence is embedded in a social environment that both shapes and is shaped by manipulation. The resulting cartography identifies cross-cutting findings, research gaps and interfaces for shared experimental settings and European capability building.

Keywords: disinformation, generative AI, influence operations, information integrity, ecosystem cartography, FIMI

1 Motivation and Lifecycle Framing

Information manipulation predates artificial intelligence. Political actors have long altered images, selectively framed events and organised distribution. The platform era industrialised personas, posts, translation and monitoring; generative AI changes the production function again. Text generation, localisation, variation and parts of persona management can now be automated at very low marginal cost. The central question is therefore not simply whether AI produces more false content, but how automation changes the relative costs and bottlenecks of an influence operation.

The international workshop *From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle*, held on 22 September 2026 at the INFORMATIK FESTIVAL in Dresden, was deliberately moderated around this problem rather than organised as independent talks. A common heuristic structured the day:

$$\textsc{Plan} \rightarrow \textsc{Target} \rightarrow \textsc{Develop} \rightarrow \textsc{Disseminate} \rightarrow \textsc{Amplify} \rightarrow \textsc{Persist/Impact}.$$

It was not intended as a deterministic sequence, but as a common language through which privacy researchers, generative-AI researchers, observatories, journalists, social scientists,

[1] IMT Atlantique, Rennes, France, marc-oliver.pahl@imt-atlantique.fr

public authorities and resilience actors could locate capabilities and dependencies. Feedback, learning and adaptation connect the stages.

A second framing hypothesis concerned cost. Schematically,

$$C_{\text{campaign}} = C_{\text{content}} + C_{\text{identity}} + C_{\text{distribution}} + C_{\text{amplification}} + C_{\text{evasion}}, \tag{1}$$

and for N items and automation share α,

$$C_{\text{content}} = N[(1-\alpha)c_{\text{human}} + \alpha c_{\text{LLM}}]. \tag{2}$$

The opening used an illustrative comparison of roughly USD 250 for a human-produced and verified 500-word article versus USD 0.015 marginal LLM generation cost, a ratio of about 16,667. This does *not* make the entire operation 16,667 times cheaper. As content cost falls, audience knowledge, credible identities, distribution, attention, trust, coordination, operational security and persuasion can become relatively scarcer. Threat-intelligence reporting supports the broader observation that AI is also entering operational functions beyond content generation [An26].

The programme progressively changed the level of analysis: journalism and verification; national ecosystems; governance and targeting; hands-on recognition and manipulation; controlled generation and propagation; human vulnerability and persuasion; and finally argumentative, evaluative and implicit manipulation. This paper reconstructs that cartography and derives connections visible only when the perspectives are considered together.

2 Methodology and Scope

This paper is an evidence-grounded workshop cartography, not a systematic literature review or representative survey of national ecosystems. Its evidence base is deliberately heterogeneous because the workshop combined scientific papers, project presentations, invited interventions, national ecosystem perspectives, hands-on activities and discussion.

For proceedings-backed contributions, the paper is the primary source: Schoo [Sc26]; Loth, Kappes and Pahl on RogueGPT [LKP26g] and the kill chain [LKP26f]; Lebon [Le26a]; Ricker et al. [Ri26]; Mortaga et al. [Mo26]; and Langner [La26]. Oral presentations are used mainly for interdisciplinary framing. For oral-only contributions—Caroline Lindekamp, Anthony Wong, the German project block, Matthieu Blondeau, Anatolii Marushchak and Renée DiResta—the transcript and speaker-provided slides constitute the primary workshop record. Limited external research explains institutions and terminology and is cited separately.

The analysis asks three questions: what capability or perspective does a contribution provide; where does it primarily intervene in the lifecycle; and which other contributions are complementary? Collaboration opportunities are prospective synthesis, not claims of existing collaborations. Participation reflects submissions, invitations, outreach and availability; absence from the map does not imply absence of a capability in the wider field.

3 Ecosystem Cartography

Table 1 condenses the substantive capabilities represented in Dresden. Defensive functions such as verification, governance and resilience are shown alongside lifecycle stages rather than treated as additional stages.

Three patterns emerge. Content analysis and detection are comparatively well represented; distribution, coordination and observability become more important as generation costs fall; and governance, trust and societal resilience shape whether technical capabilities become effective interventions.

3.1 Germany: broad research capacity

The German block was dominated by research projects, many from the BMFTR programme on detecting, understanding and countering digital disinformation [Bu26]. CleanFeed connects provenance and journalistic workflows [Fr26; Pl26]; DeMasKI combines rhetoric and interpretable AI [Un26b; Un26d]; DeMuKo studies multilingual vulnerability [DI26]; kuKI develops culturally sensitive multilingual assistance [FZ26]; SONAR extends analysis to podcasts and paralinguistic signals [Le26b]; SOURCE develops open web-scale infrastructure [Un26c]; and VITraDes studies transparency of AI use in journalism [Un26e]. MORGEN adds population-level modelling of disinformation, health effects and interventions [Un26a]; Im:DIS extends the landscape to implicit disinformation and is embedded in the DFG Priority Programme Re:DIS [De25; SLZ25]. The resulting picture is scientifically broad but institutionally distributed.

3.2 France: structured state capabilities

Blondeau highlighted a different configuration. France has developed a substantial institutional response centred on VIGINUM, the SGDSN service for detecting and characterising foreign digital interference affecting French digital public debate [SG26b]. Its focus is not generic fact-checking but coordinated modes of operation and tactics [SG26a]. The 2026–2030 national strategy emphasises resilience, platform and generative-AI responsibility, detection/attribution/response and European cooperation [Se26]. Military Lutte informatique d'influence is distinct and belongs to the operational military domain [Mi23]. The Pôle d'Excellence Cyber provides an intermediary interface across public, research and private actors. At European level, this architecture connects to the EEAS FIMI framework, Digital Services Act and European Democracy Shield [Eu24; Eu25a; Eu26]. The French perspective therefore showed comparatively mature state-operational structures while still identifying a need for scalable private capability.

Actor / contribution	Perspective	Core capability	Lifecycle / response
Lindekamp / CORRECTIV	Journalism	Verification, media literacy, trusted intermediaries	Disseminate, amplify, verify
Wong / IFIP	AI governance	Data governance, accountability, professional responsibility	Cross-cutting governance
German landscape	Research ecosystem	Provenance, language, audiences, observability, journalism, education	Across lifecycle
PEC / Blondeau	French ecosystem	FIMI, public–private capability building, coordination	Plan, coordinate, respond
Marushchak / Ukraine	Information warfare	Monitoring, authentication, communication, investigation, civil society	Full lifecycle, resilience
DiResta	Influence operations	Production–distribution–persuasion, orchestration, attention, trust	Full lifecycle, persuasion
CleanFeed	Provenance	Authenticity across media chains	Develop, disseminate, verify
DeMasKI	AI, rhetoric	Interpretable rhetorical-manipulation detection	Develop, detect, explain
DeMuKo	Audience, education	Multilingual vulnerability and trust	Target, receive, resilience
kuKI	Multilingual AI	Culturally sensitive detection/contextualisation	Develop, detect, verify
SONAR	Media analysis	Podcast/audio misinformation, paralinguistic features	Develop, disseminate, detect
SOURCE	Observability	Open web-scale campaign detection	Disseminate, amplify, persist
VITraDes	Journalism, law	Transparency of AI use in journalism	Develop, disclose, govern
MORGEN	Public health	Disinformation/impact modelling and intervention evaluation	Amplify, impact, intervene
Im:DIS	Linguistics, HCI	Presupposition, implicature and epistemic effects	Develop, persuade, explain
Schoo	Privacy, regulation	PETs, targeting economics, provable compliance	Target, govern, verify
Lebon	Social/governance	Urban crisis, vulnerability, institutional response	Amplify, impact, respond
RogueGPT / JudgeGPT	Experimentation	Controlled generation and human judgments	Develop, receive, evaluate
SPINE / OPSCI	Observability	Cross-platform narrative/coordination analysis	Disseminate, amplify, persist
Kill-chain study	Human perception	Vulnerabilities, fatigue, intervention points	Receive, intervene
Mortaga et al.	Human-centred AI	Argumentative-flaw detection/localisation	Detect, explain
Langner	Linguistics	Evaluative language and judgmental bias	Develop, persuade, detect

Tab. 1: Condensed cartography of perspectives and capabilities represented at the workshop.

3.3 Ukraine: layered resilience under information warfare

Marushchak showed the problem under conditions of ongoing war. Russian information operations have accompanied aggression against Ukraine since at least 2014 [Eu23; Eu25b; Ma26]; generative AI enters an established architecture rather than creating the underlying threat. It can assist writing, translation, persona construction, synthetic media and account management, while deepfakes create new impersonation opportunities [Ce25; MPK25].

Ukraine's response is best understood as layered resilience: monitoring, strategic communication, military authentication, investigation, civil-society fact-checking, international cooperation and platform engagement operate as complementary functions [Ce21; Ce24; Ma26]. Detection is only a first step; authentication, verified communication, investigation and intervention must follow. Wartime experience also exposes legal and jurisdictional constraints: rapid response must coexist with freedom-of-expression safeguards, while platforms outside national jurisdiction hold data unavailable to public authorities or researchers [MPK25]. Germany, France and Ukraine thus expose complementary emphases: research breadth, institutionalised operational response and layered resilience.

4 From Governance to Operational Influence

4.1 Media, trust and professional verification

Lindekamp's opening keynote situated disinformation in the environment in which citizens encounter it: journalism, platforms, fact-checking, media literacy and AI-assisted information production. Professional verification is not a binary classifier. Fact-checkers identify claims worth investigating, reconstruct context, locate primary evidence and communicate conclusions in inspectable form. This creates an asymmetry: generative systems can produce plausible content cheaply, whereas verification remains evidence-intensive.

AI-generated and false information are separate questions. JudgeGPT later operationalised this distinction, reinforcing Lindekamp's warning against treating AI origin as a proxy for falsity. Trust is equally central: users cannot verify every item, so they depend on accumulated credibility in institutions, journalists and methods. Trusted intermediaries are therefore both defensive resources and potential targets. Media literacy complements reactive debunking by strengthening upstream evaluation practices; technology can assist retrieval, clustering and evidence comparison without replacing editorial judgment.

4.2 AI governance and professional responsibility

Wong's IFIP perspective placed the lifecycle inside AI and data governance. Risks arise from combinations of models, data, targeting, accounts, platforms, recommendation systems,

organisational processes and human behaviour; governance therefore needs to address socio-technical systems rather than isolated models. Data support audience profiling, model training, targeting and evaluation, while generative systems increasingly mediate information access. This creates a direct bridge to Schoo's question of verifiable purpose limitation.

Responsibility is distributed. Model providers may see generation and orchestration but not downstream dissemination; platforms see engagement but may not know how material was produced; journalists see public claims without platform telemetry; researchers see fragments; and authorities need technically meaningful evidence for oversight. Professional communities such as GI, GChACM and IFIP can provide continuity across these communities, while Re:DIS offers an interdisciplinary scientific network [De25]. These are not lifecycle stages but institutional infrastructure for sustained capability building.

4.3 Targeting governance and socially embedded consequences

Schoo models manipulative microtargeting through two interacting engines: a *data engine* determining who is reached and a *deception engine* determining what they are told [Sc26]. He distinguishes controller-deployed PETs from user-controlled tools whose population-scale effect depends on adoption. PETs primarily constrain TARGET; they do not prevent generic synthetic deception at DEVELOP or DISSEMINATE. Privacy can also become an experimental intervention: degrading audience data can test how targeting accuracy and campaign economics change.

The strongest bridge to governance is the demand that compliance become *provable rather than promised*. Provenance, usage logs, cryptographic auditing and machine-readable policies could make selected data-use obligations independently auditable. For targeted communication, provenance can extend beyond "where did this artefact come from?" to "was the audience data used for an authorised purpose?" [Sc26].

Lebon's North-Holland case provides the complementary social perspective [Le26a]. Manipulated information does not originate in a self-contained online sphere and then enter the "real world"; narratives draw from existing events, tensions and vulnerabilities from the outset. In the case studied, schools closed, authorities communicated under uncertainty and fear persisted. Social context is both an input to manipulation and an outcome potentially reshaped by circulation. Together, Schoo and Lebon suggest defence in depth across data governance, information flows, institutions and social context.

4.4 DiResta: production, distribution and persuasion

DiResta's invited intervention connected the workshop's technical and human perspectives by separating *production*, *distribution* and *persuasion*. Generative AI clearly lowers production costs and can support translation, localisation and persona maintenance. Increasingly,

however, AI also enters the operational "back office": audience research, intelligence processing, databases, persona histories, coding, timing, quality control and workflow orchestration. Provider reporting documents analogous uses beyond visible content generation [An26].

This matters because agentic workflows can retain state, use tools and connect sequences of tasks, potentially integrating audience analysis, content production, persona management and feedback. Yet cheap production does not solve distribution. High-volume networks can generate abundant content without attracting authentic engagement. Attention, credible identities and trusted messengers may therefore become more valuable as synthetic content becomes cheaper.

Persuasion remains a distinct and weakly observed outcome. Reach, impressions and engagement do not establish changes in belief, trust or behaviour. Conversational systems may create new persuasive settings through sustained interaction, while attempts to manipulate retrieval sources can move influence upstream into AI-mediated information environments; source-skew research illustrates why retrieval itself matters [KD26]. A further observability gap follows: model providers may see planning and generation, platforms see dissemination, and neither alone sees the complete operation. DiResta's central contribution to the synthesis is therefore a constraint on technological determinism: AI can transform production and orchestration without automatically producing attention or persuasion.

5 Generation, Propagation and Human Vulnerability

5.1 Controlled generation as research infrastructure

RogueGPT moves from ad-hoc prompting toward controlled experimental infrastructure [LKP26g; LKP26h]. It supports configurable models, languages, personas, narrative framing and output styles so that synthetic misinformation becomes a reproducible experimental variable rather than an uncontrolled artefact. JudgeGPT provides the complementary downstream instrument by separating perceived origin from perceived veracity. Earlier experiments found that participants could not reliably distinguish LLM-generated from human-written news and that performance can deteriorate during sustained evaluation [LKP26a]. The two systems have already been framed as an experimental pipeline for studying LLM-generated misinformation [LKP26e].

The Dresden RogueGPT paper extends this foundation toward agentic research infrastructure: controlled stimuli can seed simulations in which personas, narratives and downstream behaviour are studied across a wider operation [LKP26g]. This creates an interface with observability rather than an argument for autonomous real-world dissemination.

5.2 SPINE: observing propagation

Ricker et al. move downstream with SPINE, the third iteration of the European Narrative Observatory [Ri26]. It combines narratology, semantic analysis and coordination detection, with explicit mapping of national digital public spheres and canonical claim construction. The Romanian case illustrates differentiated platform roles: Telegram as a seeding or command layer, TikTok as a virality engine and Facebook as memory infrastructure in which narratives persist and resurface. Coordination is treated as a signal rather than automatic proof of disinformation, and LLMs assist analyst adjudication rather than replacing it end-to-end.

RogueGPT and SPINE are therefore complementary: one can provide controlled cross-lingual seeds and simulation inputs; the other provides propagation and coordination measurements. SOURCE and MORGEN extend this pipeline toward web-scale observation and population-level consequences.

5.3 Human vulnerability and intervention

The kill-chain study uses the chain as an intervention taxonomy rather than a deterministic process model [LKP26f]. It maps 2,438 JudgeGPT judgments from 504 participants onto stage-specific vulnerabilities: accuracy was 58.4% for identifying AI-generated content and 68.1% for fake-content detection. Fast judgments performed worse than deliberate ones, while sustained exposure reduced fake-news detection by about ten percentage points. Earlier JudgeGPT work similarly found difficulty distinguishing LLM-generated from human-written news [LKP26a]; causal analysis further points to expertise, familiarity and cognitive fatigue as relevant dimensions [LKP26c].

These results motivate layered interventions: provenance where human origin detection fails, friction or deliberation prompts at delivery, and cognitive scaffolding against fatigue. More broadly, they show why measuring an operation requires more than counting produced artefacts or impressions: the eventual question is how exposure changes judgment and behaviour.

6 Manipulation Beyond Factual Falsity

Mortaga, Sieger, Goeke and Calero Valdez argue that binary misinformation detection is insufficient for decision support: users need to understand *what* makes a post misleading [Mo26]. Their study operationalises argumentative structure as claim, ground and warrant and compares four LLMs with a 120-participant human baseline. All tested LLMs exceeded the human binary-detection baseline of 63%, with the best configuration reaching 91%, but localisation remained substantially harder. The implication is content-adaptive human–AI decision support, not a universal automated truth oracle.

Langner exposes a complementary blind spot: manipulation can occur even when propositional content is factually correct [La26]. Scalar evaluative expressions add an attitudinal layer that can shape interpretation, particularly when recipients lack domain expertise. The proposed denial-of-expectation approach compares observed values with expected values or data-derived thresholds to assess whether an evaluation is licensed by the underlying distribution.

Im:DIS extends this problem below explicit assertion. Presuppositions, implicatures and related pragmatic inferences can introduce contestable assumptions into common ground without stating them directly. Existing experiments show that LLMs can struggle to reject false information introduced through presuppositions [SLZ25]. Together, Mortaga et al., Langner and Im:DIS broaden information integrity from truth classification toward reasoning, framing, pragmatic inference and explanation.

7 Cross-Cutting Research Agenda

The cartography reveals a research agenda that extends beyond detection of individual false or AI-generated artefacts. Rather than retain thirteen separate challenges from the full workshop report, we condense them into six interfaces where represented capabilities can be combined.

7.1 Operation-level observability

Influence operations connect planning, audience selection, generation, dissemination, amplification, persistence and adaptation. SPINE demonstrates semantic, temporal and coordination analysis [Ri26]; SOURCE points toward web-scale infrastructure; model-provider reporting illustrates visibility at the generation layer [An26]. Yet no actor sees the whole chain. The challenge is to combine evidence across providers, platforms, open-source observation and institutional actors while preserving privacy, security and legitimate access restrictions.

7.2 Causal measurement of persuasion and sustained exposure

Production, distribution and persuasion must be measured separately. Reach and engagement establish exposure, not belief change. Human performance also changes with deliberation and sustained exposure [LKP26c; LKP26f]. Research therefore needs longitudinal and ethically feasible methods that measure belief, trust and behaviour, including messenger credibility, repetition, personalisation and conversational interaction.

7.3 Economics and automation of influence operations

Schoo's two-engine model makes the cost-shift hypothesis experimentally tractable: controlled generation can vary the information flow, while PETs can deliberately degrade the audience-data flow [LKP26g; Sc26]. Agentic systems add a further question: which operational functions can realistically be automated, how does automation change observable coordination signatures, and which bottlenecks remain scarce? Controlled simulation, observability and population-level models such as MORGEN offer complementary measurement layers.

7.4 Trustworthy information infrastructure

CleanFeed addresses artefact provenance, VITraDes the communication of AI involvement, and Schoo purpose-compliant use of audience data. Expert studies also identify limitations of content detection and motivate reproducible provenance [LKP26i]. The challenge is to connect provenance, secure logging, cryptographic auditing and machine-readable policies without creating new surveillance risks. As AI systems mediate information access, source quality and retrieval manipulation also become integrity problems [KD26].

7.5 Explainable human–AI decision support

AI support should not merely replace human judgment with another binary classifier. Argument localisation [Mo26], evaluative-language analysis [La26], implicit-disinformation research and source-level credibility signals provide complementary evidence layers. CRED-1, for example, develops a multi-signal domain credibility resource for privacy-preserving pre-bunking [LKP26b]. The design question becomes which information should reach humans at which point: provenance before exposure, source context at delivery, claim- and argument-level evidence during evaluation, and professional verification when uncertainty remains.

7.6 From social context to operational resilience

Lebon's case shows that social context is simultaneously input and outcome: events and vulnerabilities shape narratives, circulation affects institutional behaviour, and those reactions can generate new material for manipulation [Le26a]. National comparison adds a translation problem: research methods do not automatically become operational capacity. Domain also matters; expert assessment finds different threat profiles and countermeasure preferences for political and health disinformation [LKP26d]. The challenge is therefore to build shared infrastructures and governance arrangements that turn research capabilities

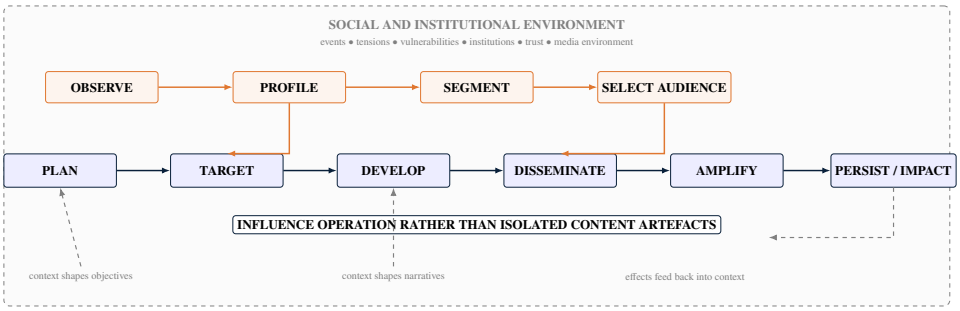


Fig. 1: Refined socio-technical model emerging from the workshop synthesis. Information and audience-data flows interact inside a social and institutional environment; feedback and adaptation make the system iterative rather than linear.

into context-sensitive operational resilience while preserving democratic oversight and fundamental rights.

These interfaces suggest shared experimental settings rather than isolated projects. A lifecycle-wide setting could connect controlled generation and targeting, propagation and cross-platform observation, repeated human exposure, measures of belief and trust, provenance and compliance mechanisms, and institutional response. RogueGPT/JudgeGPT, SPINE/SOURCE, MORGEN, professional fact-checking and public-sector environments already provide components of such a pipeline.

8 Synthesis: An Iterative Socio-Technical Model

The most important outcome of Dresden is a change in the unit and structure of analysis. The workshop began with a deliberately simple lifecycle and progressively exposed three refinements. First, the unit of analysis expands from individual content artefacts to the wider influence operation. Second, Schoo's two-engine perspective reveals interacting *information* and *audience-data* flows rather than a single chain [Sc26]. Third, Lebon's analysis shows that the operation is embedded in a social environment rather than moving from an autonomous online sphere into the "real world" [Le26a]. Existing events and vulnerabilities shape narratives and targeting, while circulation can feed back into perceptions, institutions and social conditions.

Eight cross-cutting findings summarise the synthesis:

1. **Analysis must expand from content to operations.** Artefact detection cannot substitute for understanding planning, targeting, distribution and adaptation [LKP26g; Ri26].

2. **Generative AI shifts bottlenecks.** Production becomes cheaper while audience access, credible identities, distribution, attention, trust and persuasive relationships can remain scarce.
3. **Automation reaches the operational back office.** Audience analysis, persona management, coding and workflow orchestration can also be AI-assisted [An26].
4. **Manipulation is embedded in real-world context.** Existing tensions shape narratives and circulation can feed back into institutions and trust [Le26a].
5. **Technical and regulatory responses are complementary.** Provenance, privacy and detection require governance, while governance increasingly requires technically meaningful evidence [Sc26].
6. **Trust and credible distribution are strategic resources.** Cheap synthetic content does not eliminate dependence on trusted messengers and durable audience relationships.
7. **No actor sees the complete lifecycle.** Providers, platforms, researchers, journalists and authorities possess different fragments of evidence [An26; Ri26].
8. **Manipulation can be factually and pragmatically subtle.** Argumentative flaws, evaluative language, presuppositions and implicatures require analysis beyond factual verification [La26; Mo26; SLZ25].

9 Conclusion and Continuity

The workshop suggests that many capabilities required to study and counter AI-enabled information manipulation already exist but remain distributed across disciplines, institutions and national ecosystems. Germany contributes substantial research breadth; France illustrates structured state-operational capabilities; Ukraine demonstrates layered resilience under ongoing information warfare; journalism contributes verification practice and trust; and professional communities can provide institutional continuity.

The next step should therefore not simply be another detector or thirteen isolated research projects. It should be shared experimental and operational settings in which controlled generation, audience data, propagation, human response, provenance, governance and institutional intervention can be studied together. Such settings would allow the workshop's central hypotheses—cost shifts, agentic orchestration, persuasion, observability and socially embedded feedback—to become empirically testable.

The present cartography is exploratory. A subsequent expert-validation phase can prioritise the findings, identify missing or overstated issues and test where interdisciplinary collaboration is most necessary. The longer-term objective is a living capability map and a sustained research and operational environment for disinformation, generative AI and information integrity. The principal result of Dresden is that many of the pieces are already visible; the next task is to connect them and evaluate how they work together.

Full report and living resource centre. This proceedings contribution presents a condensed synthesis of the workshop. The open-access workshop report provides detailed documentation of the invited and oral-only interventions, national ecosystem perspectives, contribution cartography, thirteen research challenges and collaboration interfaces [Pa26]. The full report is available at https://doi.org/10.5281/zenodo.23451417. The accompanying *Disinformation Space* resource centre at https://disinfo.spy360.net/ connects the citable workshop outcomes to an evolving research-project and capability map, publications and opportunities for interdisciplinary collaboration.

References

[An26] Anthropic: Detecting and Countering Misuse of AI: September 2026, https://www.anthropic.com/threat-intelligence-report-september-2026, Threat Intelligence Report, 2026.

[Bu26] Bundesministerium für Forschung, Technologie und Raumfahrt: Desinformation – Erkennen. Verstehen. Abwehren.: Neue Forschungsprojekte gestartet, BMFTR funding initiative; eleven research projects funded from 2026 to 2029, 2026.

[Ce21] Center for Countering Disinformation, National Security and Defense Council of Ukraine: About Center for Countering Disinformation, https://cpd.gov.ua/en/docs/about-center-for-countering-disinformartion/, Official institutional description, 2021.

[Ce24] Center for Countering Disinformation, National Security and Defense Council of Ukraine: The Center Held a Press Conference on the Results of Its Work, https://cpd.gov.ua/en/events-en/the-center-held-a-press-conference-on-the-results-of-its-work/, Official description of monitoring and inter-agency/platform cooperation, 2024.

[Ce25] Center for Countering Disinformation, National Security and Defense Council of Ukraine: Analytical Report: The Use of Artificial Intelligence-Generated Videos to Discredit the Defence Forces of Ukraine, https://cpd.gov.ua/en/report/analytical-report-the-use-of-artificial-intelligence-generated-videos-to-discredit-the-defence-forces-of-ukraine/, Documents the 16 March 2022 Zelenskyy surrender deepfake and later AI-generated video operations, 2025.

[De25] Deutsche Forschungsgemeinschaft: Priority Programme "Rethinking Disinformation (Re:DIS)" (SPP 2573), https://www.dfg.de/de/aktuelles/neuigkeiten-themen/info-wissenschaft/2025/ifw-25-45, Information für die Wissenschaft Nr. 45, 2025.

[DI26] DIPF – Leibniz Institute for Research and Information in Education: DeMuKo – Selbstbestimmt und informiert: Desinformation im multilingualen Kontext verstehen und vorbeugen, https://www.dipf.de/de/forschung/projekte/demuko-selbstbestimmt-und-informiert-desinformation-im-multilingualen-kontext-verstehen-und-vorbeugen-teilprojekt-dipf, Project description, accessed September 2026, 2026.

[Eu23] European Union: EU Statement on the Russian Federation's Ongoing Aggression against Ukraine, https://www.eeas.europa.eu/delegations/vienna-international-organisations/osce-permanent-council-no1452-vienna-23-november-2023_en, Discusses Russian information manipulation against Ukraine since 2014, 2023.

[Eu24] European External Action Service: 2nd EEAS Report on Foreign Information Manipulation and Interference Threats, https://www.eeas.europa.eu/eeas/2nd-eeas-report-foreign-information-manipulation-and-interference-threats_en, FIMI threat and response framework, 2024.

[Eu25a] European Commission: European Democracy Shield, https://commission.europa.eu/publications/european-democracy-shield-documents_en, Official European Commission documents, 2025.

[Eu25b] European External Action Service: Report on FIMI Threats, https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf, Third EEAS FIMI threat report, 2025.

[Eu26] European Commission: The Digital Services Act, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act, Official overview of systemic-risk and platform-accountability obligations, 2026.

[Fr26] Fraunhofer Institute for Open Communication Systems FOKUS: CleanFeed: Trust in Digital Media, https://www.fokus.fraunhofer.de/en/fame/projects/cleanfeed.html, Project description, accessed September 2026, 2026.

[FZ26] FZI Research Center for Information Technology: Fact-Checking across Language Barriers: AI Assistant Detects Disinformation in German, Russian, and Turkish, https://www.fzi.de/en/2026/04/23/kuki-against-disinformation/, Project announcement, 2026.

[KD26] Klonick, K.; DiResta, R.: Invisible Source Skew. Lawfare, 2026.

[La26] Langner, M.: Scalar Evaluative Expressions and Their Manipulative Use in LLM-Generated Text. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[Le26a] Lebon, L.: AI-Generated Disinformation in Times of Urban Crisis: Lessons from the Rival Youth Groups Case in North Holland. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[Le26b] Leibniz Institute for Media Research | Hans-Bredow-Institut: Podcasts: Actors, Topics, and Misinformation, https://leibniz-hbi.de/en/hbi-projects/podcasts-actors-content-and-misinformation/, SONAR project description, accessed September 2026, 2026.

[LKP26a] Loth, A.; Kappes, M.; Pahl, M.-O.: Can Humans Tell? A Dual-Axis Study of Human Perception of LLM-Generated News. In: Companion Publication of the 2026 18th ACM Web Science Conference. ACM, pp. 34–39, 2026, https://doi.org/10.1145/3795513.3807431.

[LKP26b] Loth, A.; Kappes, M.; Pahl, M.-O.: CRED-1: An Open Multi-Signal Domain Credibility Dataset for Automated Pre-Bunking of Online Misinformation, Preprint, 2026, https://doi.org/10.2139/ssrn.6448466.

[LKP26c] Loth, A.; Kappes, M.; Pahl, M.-O.: Eroding the Truth-Default: A Causal Analysis of Human Susceptibility to Foundation Model Hallucinations and Disinformation in the Wild. In: Companion Proceedings of the ACM Web Conference 2026. ACM, pp. 1125–1132, 2026, https://doi.org/10.1145/3774905.3795832.

[LKP26d] Loth, A.; Kappes, M.; Pahl, M.-O.: Experts Disagree on How to Fight AI Disinformation, but Agree That Health and Politics Need Different Solutions. Harvard Kennedy School Misinformation Review 7 (4), 2026, https://doi.org/10.37016/mr-2020-205.

[LKP26e] Loth, A.; Kappes, M.; Pahl, M.-O.: Industrialized Deception: The Collateral Effects of LLM-Generated Misinformation on Digital Ecosystems. In: Companion Proceedings of the ACM Web Conference 2026. ACM, pp. 295–302, 2026, https://doi.org/10.1145/3774905.3795471.

[LKP26f] Loth, A.; Kappes, M.; Pahl, M.-O.: Interrupting the Chain: Human Perception of AI-Generated Disinformation Through a Kill Chain Lens. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[LKP26g] Loth, A.; Kappes, M.; Pahl, M.-O.: RogueGPT as Agentic Infrastructure: Seeding Cross-Lingual Misinformation Simulations and Red-Teaming Social Media Agents. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[LKP26h] Loth, A.; Kappes, M.; Pahl, M.-O.: RogueGPT: A Controlled Stimulus Generation Framework for News Authenticity Research. Journal of Open Source Software 11 (125), p. 11219, 2026, https://doi.org/10.21105/joss.11219.

[LKP26i] Loth, A.; Kappes, M.; Pahl, M.-O.: The Verification Crisis: Expert Perceptions of GenAI Disinformation and the Case for Reproducible Provenance. In: Companion Proceedings of the ACM Web Conference 2026. ACM, pp. 980–988, 2026, https://doi.org/10.1145/3774905.3795484.

[Ma26] Marushchak, A. et al.: Counter Disinformation System Development in Ukraine. Springer, Cham, 2026.

[Mi23] Ministère des Armées et des Anciens combattants: La lutte informatique d'influence (L2I), https://www.defense.gouv.fr/comcyber/nos-operations/lutte-informatique-dinfluence-l2i, Official COMCYBER description, 2023.

[Mo26] Mortaga, M. et al.: Not All Errors are Equal: Comparing Humans and LLMs on Detecting Argumentative Flaws in Misinformation. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[MPK25] Marushchak, A.; Petrov, S.; Khoperiya, A.: Countering AI-Powered Disinformation through National Regulation: Learning from the Case of Ukraine. Frontiers in Artificial Intelligence 7, p. 1474034, 2025, https://doi.org/10.3389/frai.2024.1474034.

[Pa26] Pahl, M.-O.: From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle — Workshop Report, Ecosystem Cartography and Research Agenda, Workshop report, version 1.0, Zenodo, 2026, https://doi.org/10.5281/zenodo.23451417.

[Pl26] Plaum, A.: CleanFeed: Verified, Transparent Journalism in An Era of Disinfo and AI Slop, https://innovation.dw.com/articles/cleanfeed-verified-transparent-journalism-in-an-era-of-disinfo-and-ai-slop, Deutsche Welle Innovation, 2026.

[Ri26] Ricker, J. et al.: From Peripheral Seeding to Systemic Contamination: Multi-Vector Monitoring of Information Propagation. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[Sc26] Schoo, P.: More Regulation, Please: Technical Privacy Protection Alone Can Only Partially Limit Democratic Harm from Microtargeting – A Position Statement. In: Proceedings of the INFORMATIK FESTIVAL 2026 Workshop "From Prompt to Propaganda: Generative AI and the New Disinformation Lifecycle". Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V., 2026.

[Se26] Secrétariat général de la défense et de la sécurité nationale: Stratégie nationale de lutte contre les manipulations de l'information 2026–2030, https://www.sgdsn.gouv.fr/publications/strategie-nationale-de-lutte-contre-les-manipulations-de-linformation-2026-2030, Official French national strategy, 2026.

[SG26a] SGDSN – VIGINUM: G7 RRM – Adoption d'un cadre commun sur le concept de "Mode opératoire informationnel", https://www.sgdsn.gouv.fr/viginum/actualites/g7-rrm-adoption-dun-cadre-commun-sur-le-concept-de-mode-operatoire, Official French government source, 2026.

[SG26b] SGDSN – VIGINUM: Les missions de VIGINUM, https://www.sgdsn.gouv.fr/viginum/comprendre-viginum/les-missions-de-viginum, Official French government source, 2026.

[SLZ25] Sieker, J.; Lachenmaier, C.; Zarrieß, S.: LLMs Struggle to Reject False Presuppositions when Misinformation Stakes are High. In: Proceedings of the Annual Meeting of the Cognitive Science Society. Vol. 47, 2025.

[Un26a] Universität zu Lübeck, Institut für Menschzentrierte Interaktive Systeme: MORGEN – Modellierung für Offene Resiliente Gesellschaften, https://www.imis.uni-luebeck.de/de/forschung/projekte/morgen, Project description, accessed September 2026, 2026.

[Un26b] University of Bamberg: BMFTR Project DeMasKI, https://www.uni-bamberg.de/en/cogsys/research/projects/bmftr-project-demaski/, Project description, accessed September 2026, 2026.

[Un26c] University of Passau: BMFTR Project SOURCE: European Research Infrastructure to Combat Disinformation Campaigns on the Web, https://www.uni-passau.de/en/research/research-projects-and-special-funding/project-details/research_project/source-1, Project description, accessed September 2026, 2026.

[Un26d] University of Technology Nuremberg: DeMasKI – Disinformation and Its Patterns, https://www.utn.de/en/departments/department-liberal-arts-and-social-sciences/demaski-disinformation-and-its-patterns/, Project description, accessed September 2026, 2026.

[Un26e] University of Tübingen: Trustworthy Information and Transparency Regarding Algorithms and Machine Learning Systems in Journalism (VITraDes), https://uni-tuebingen.de/en/296025, Project description, 2026.