Definitions of key terms from the presentation by Prof. Dr. Marten Risius

Definitions based on information from Google AI and ChatGPT

Echo chambers
Digital spaces (e.g., social media groups) in which people almost exclusively see information that confirms their own beliefs, which can amplify misinformation. Echo chambers consist only of people with the same views, which deprives them of exchange with the outside.

Doxing
Publication of a person's private information (e.g., address, phone number) on the internet, often as a strategy of intimidation or punishment.

Digital vigilantism
Self-administered 'justice' on the internet—when users punish people on their own (e.g., via pile-ons, shaming, or hate), instead of involving legitimate institutions.

Organized ransomware
Attacks by hacker groups that encrypt data and demand ransom—also used to sabotage media and information channels.

Shadow banning
Invisible restriction of accounts or posts without the affected users noticing (e.g., reduced reach) in order to curb misinformation.

Chatbot counter-narratives
AI-powered chatbots that respond specifically to false information by injecting counter-arguments or fact-checks.

LLM-based detoxification
Use of large language models (LLMs) to detect toxic or manipulative content and reduce it, or to rewrite it into more neutral language.

Prebunking
Preventive education that informs people in advance about typical fake-news strategies so they can better recognize manipulation.

Fiduciary AI agents
AI systems intended to act in users' best interests (e.g., filters against disinformation), similar to a 'digital fiduciary.'

Red herring
A distraction in argumentation—an irrelevant topic is introduced to divert attention from the actual fake news or critical questions.

Cherry-picking
Selective choosing: deliberately taking the best-fitting items from a set while ignoring the rest.

False dichotomies
Logical fallacies in which a situation is falsely presented as a choice between only two extreme, mutually exclusive options, although more possibilities and shades of gray exist.

Ad hominem
A rhetorical tactic or fallacy in which the person is attacked instead of the issue, to weaken the opponent's arguments and influence the audience.

Polarization
In political contexts, either a social differentiation that leads to controversy or a strengthening of differences of opinion. Often both are connected.

Imitation
Imitating or copying the behavior, style, or content of others.

Slippery-slope arguments (Dammbruch-Argumente)
Rhetorical arguments claiming that a seemingly harmless initial action (A) inevitably leads to a chain of negative consequences that culminate in a morally unacceptable end state (B).

Decontextualization
Deliberately removing information, texts, images, or videos from their original context to give them a new, often misleading meaning.

Troll
A person—or sometimes a program (bot)—that deliberately posts provocative, offensive, or misleading content to disrupt discussions and provoke or harass other users.

Bots
Automated computer programs that independently perform repetitive tasks faster than a human.

Cloning
Creation of identical copies.

Domains
A domain is a human-readable, unique name for an area on the internet that serves as an address to locate a specific website, instead of having to remember a hard-to-remember numeric string (IP address).

Gatekeeper
A gatekeeper is a person who controls access to something (e.g., a bouncer at a club door) or, more abstractly, controls who gains access to a category or status.