

Introduction

Focus on how genAI technologies (e.g., ChatGPT, DeepSeek, Gemini) impact consumer behavior, trust, and decision-making, particularly in the spread of mis, dis, and mal-information (MDM).

Proposed Solution The study provides a systematic analysis of global regulatory and policy frameworks as well as AI tools to address MDM risks and optimize the interplay between humans and genAI moderation.

Key objectives

- Map existing regulatory frameworks to limit the spread of MDM and its impact on digital platform consumers to foster digital resilience.
- To examine governance frameworks' success, adaptability, and coherence to address genAI-enhanced MDM risks.
- To explore AI detection tools and political risks and balance innovation with societal protection.?
- Identify gaps in regulatory coherence and standardization for cross-border data, international cooperation, and AI ethics.
- To recommend policy-making options to address societal and technological contexts of regions.

Research Questions (RQs)

- How do legal frameworks and AI tools address MDM challenges and cognitive biases?
- How effective are current regulatory frameworks in tackling MDM challenges?
- What strategies/ policies could help to enhance digital resilience against MDM?
- What gaps exist in genAI regulatory methods, and how can policy-making achieve global harmonization for societal benefit?

Research Methodology

Approach Systematic review and desk study focusing on global regulatory frameworks and consumer behaviour.

Systematic review PRISMA guidelines were followed to ensure rigor and transparency.

Inclusion Criteria Peer-reviewed articles, white papers, and regulatory frameworks addressing MDM and digital resilience.

Analysis Comparative analysis of regulatory frameworks, AI tools, and policy gaps.

Data Sources Scopus, Web of Science, government reports, and regulatory guidelines.

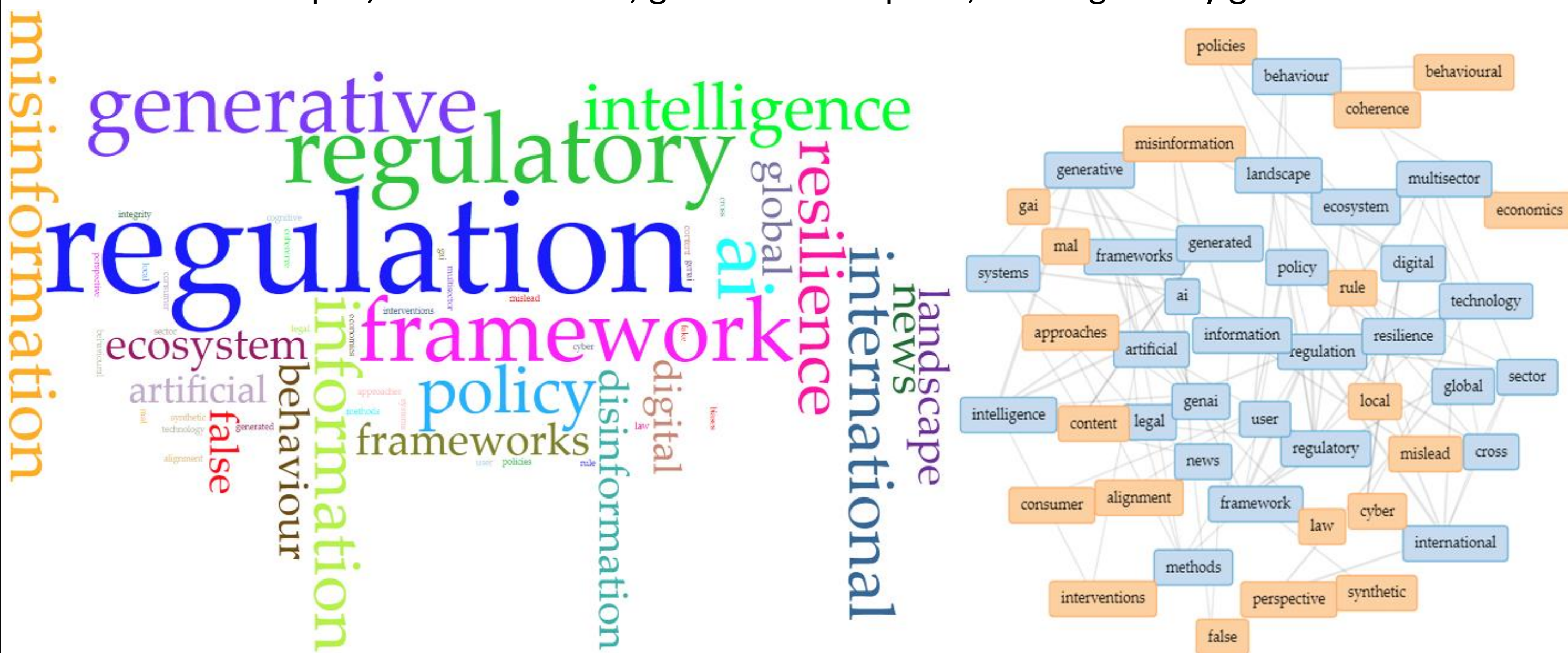


Fig. 1 Visualization chart of the search string produced through text analysis using the VOYANT tool

- Quality assessment Pico portal for evaluation of selected studies.

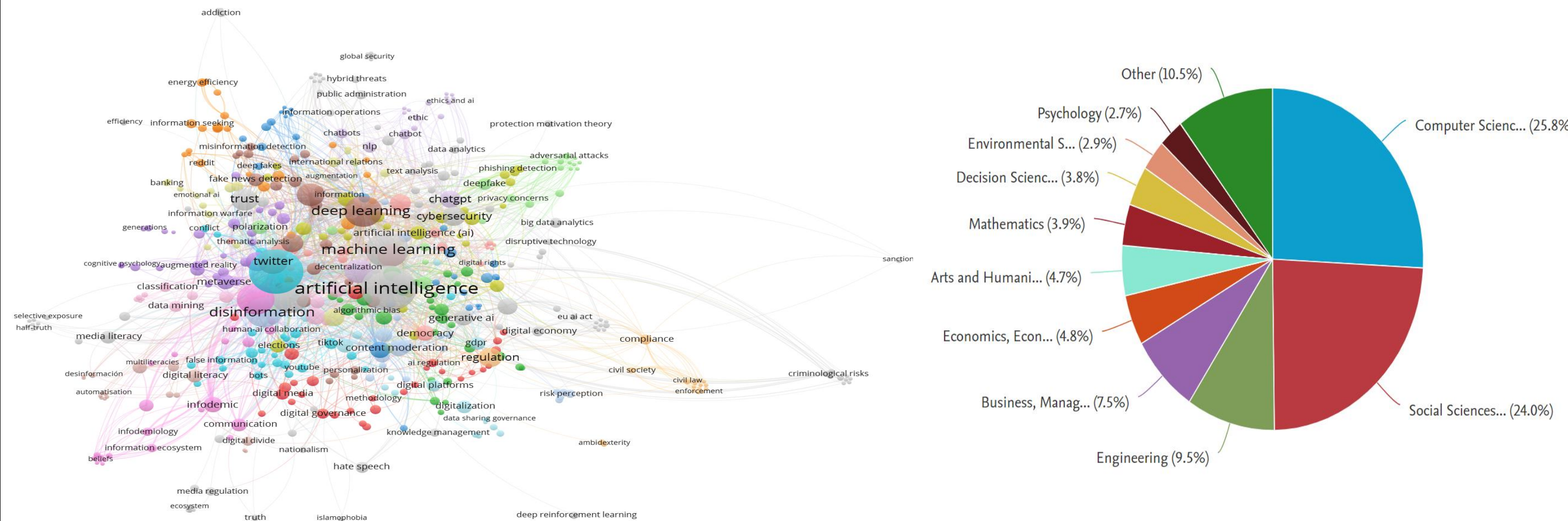


Fig 2 Keyword co-occurrence analysis of studies with VOSviewer software

Fig. 3 Distribution of studies based on the search string by research area

Results and Analysis

Key Findings

- Identifies key policy gaps in genAI regulations globally.
- Explores genAI detection tools (GBERT, Grover, SIDE) and political risks and balances innovation with societal protection.
- Examines the role of cognitive biases and MDM amplification on digital platforms.
- Shows how technical and regulatory tools can be integrated for robust content governance.
- Offers actionable policy options to boost global coherence and digital resilience.

Challenges & Policy Implications

- Regulatory Inconsistencies:** Fragmented governance across jurisdictions affects global coherence.
- AI-Driven MDM Risks:** Lack of effective detection and enforcement mechanisms.
- Cognitive Biases:** Existing policies overlook behavioral science in addressing MDM.
- Technical Challenges:** Fact-checking AI vs. adversarial manipulation.
- Enforcement & Compliance:** Weak enforcement of AI-generated content regulations.

Cognitive biases and behavioral science for Digital Resilience

- Gaps:** Current frameworks overlook psychological factors shaping MDM consumption and trust.
- Echo chambers:** Pre-existing beliefs and digital media algorithms reinforce biased information consumption.
- Behavioral nudges:** Platforms should prompt users to verify content to reduce bias-driven MDM sharing.
- Confirmation bias:** Users favor information aligning with beliefs and undermine content moderation efforts.
- Behavioral science integration:** Future regulations should incorporate insights to enhance digital resilience against MDM.
- Regulatory harmonization:** Policies should address cognitive biases to effectively combat MDM.

Challenges and Pathways to Digital Resilience

- Echo tamper-resistant watermarks:** Embed metadata in AI-generated content for origin tracing and accountability.
- Mandatory AI content labeling:** Enhances transparency and user awareness of synthetic media (China's AI Regulations, 2024).
- Cross-border policy harmonization:** EU's DSA and DMA promote global cooperation in combating MDM.
- Pre-bunking strategies:** Controlled exposure to misinformation equips users to recognize and resist MDM.
- Digital literacy initiatives:** Improve public awareness and robustness against MDM campaigns.
- Real-time flagging:** AI systems reduce reliance on human moderators and scale detection efforts.
- Ethical governance:** Balances innovation with societal concerns, addressing biases and discrimination.
- Global regulatory frameworks:** Address cross-border MDM challenges to ensure uniform compliance.

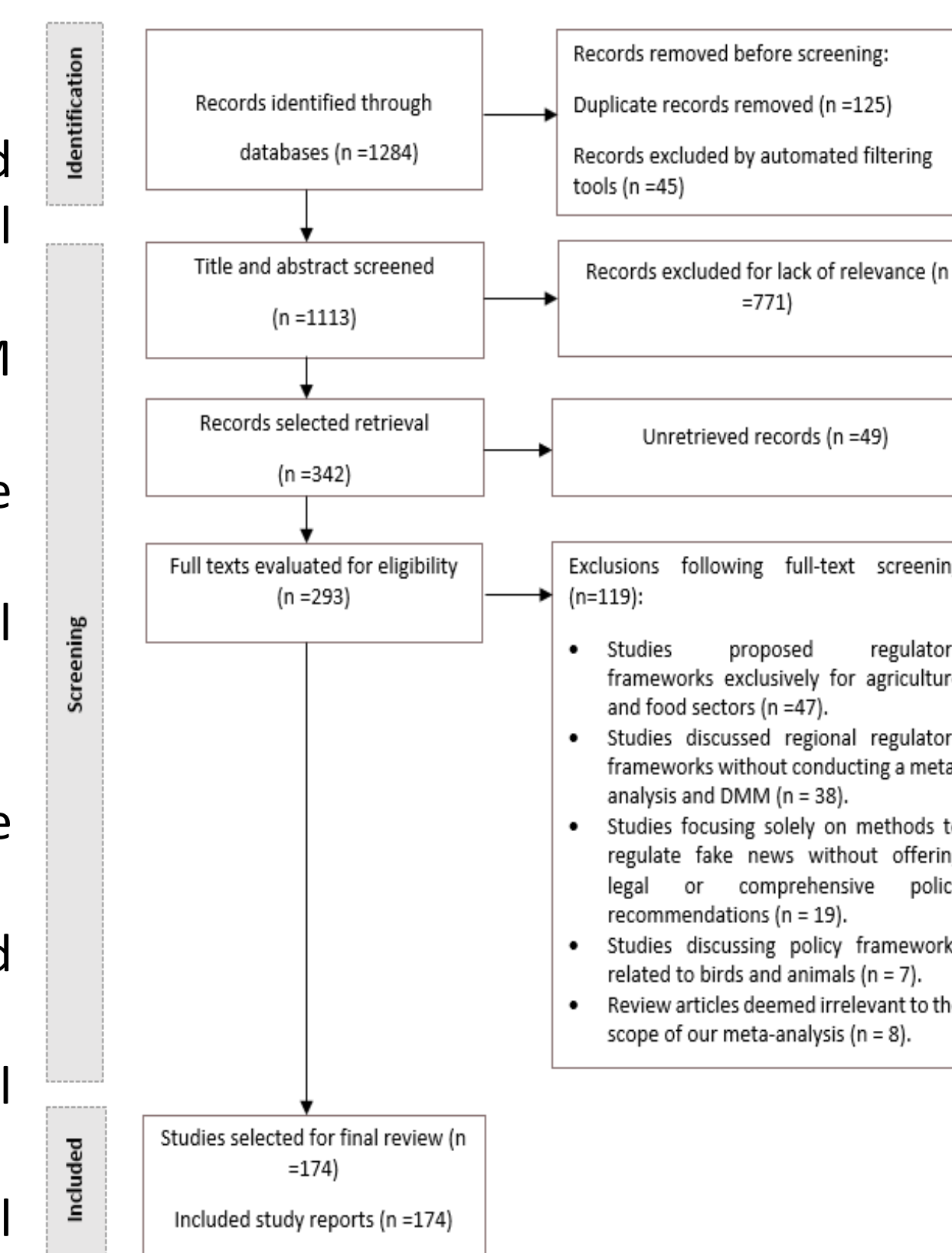


Figure. 4 PRISMA Flow Diagram: Overview of data selection and screening in systematic reviews process

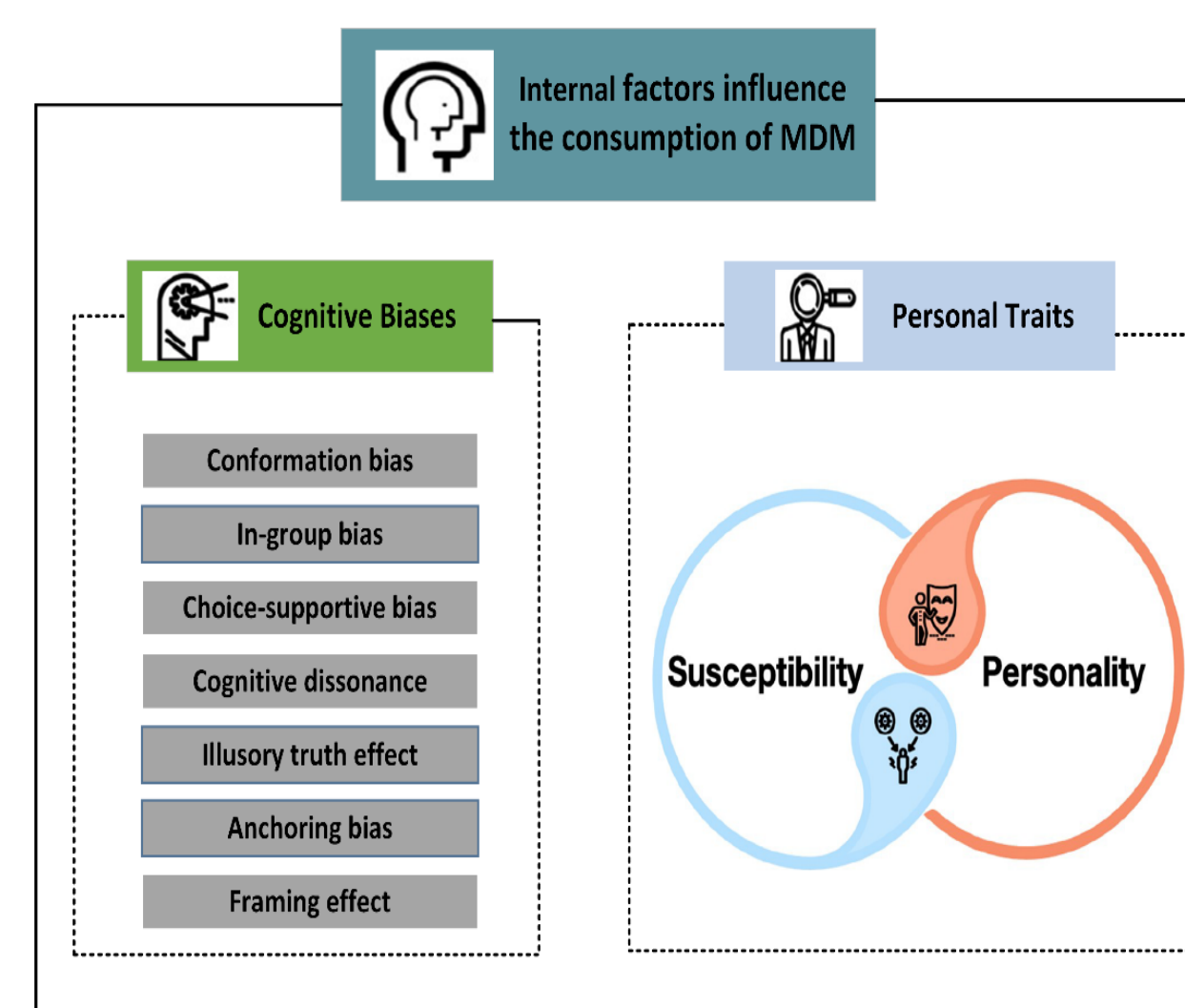
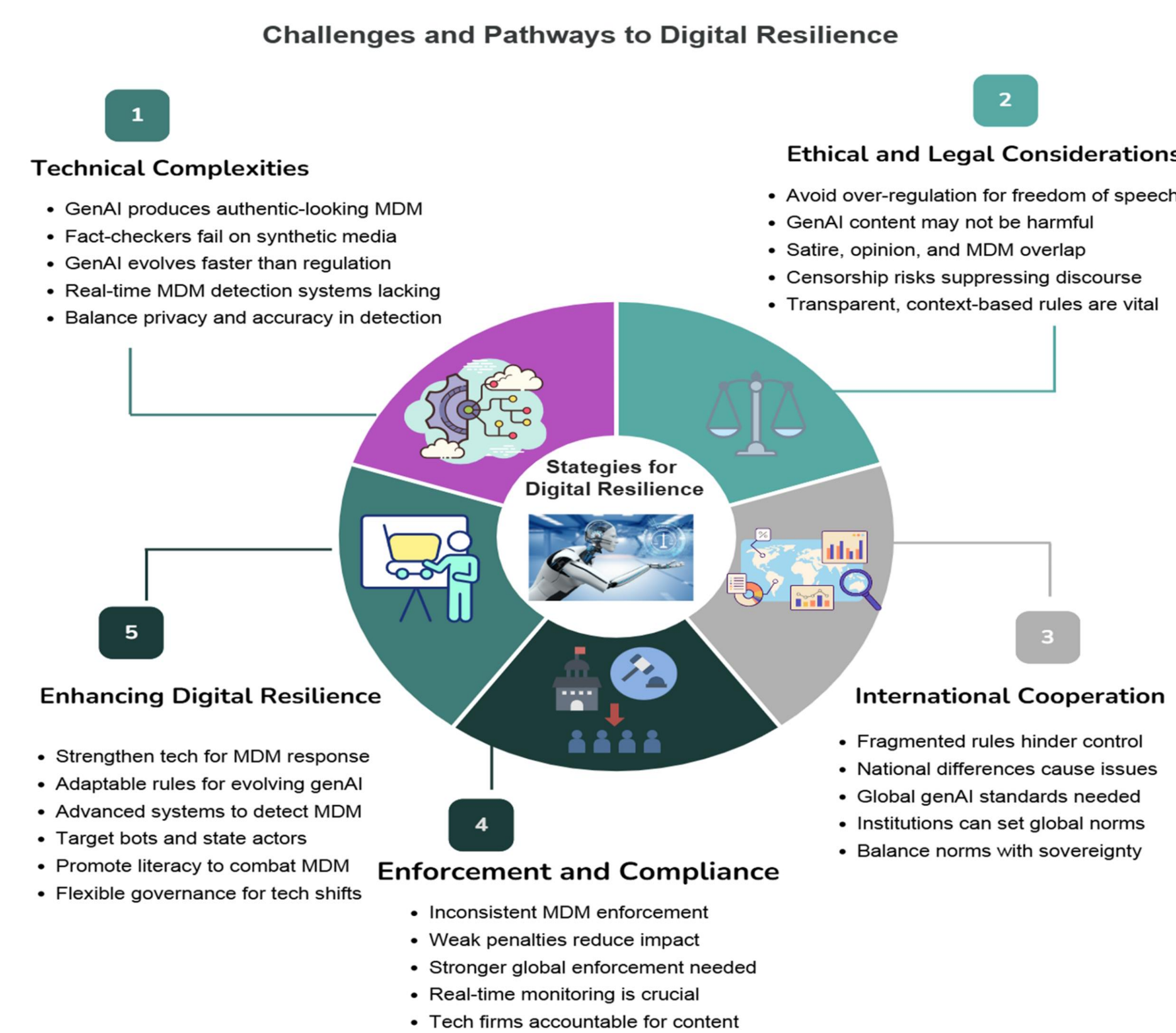


Fig. 5 Internal factors influence the consumption of MDM



Desk Study & Policy-making

Desk Study outcomes

Mapping of global regulatory frameworks

- Risk-based regulation:** Prioritizes high-harm MDM risks, enabling tailored interventions for public safety and trust.
- Rule-based regulation:** Enforce clear, consistent standards but lack flexibility for evolving genAI-driven MDM challenges.
- Self-regulation:** Empowers tech companies to adopt ethical guidelines, reducing regulatory burden. However, it faces challenges due to conflicts of interest and public trust.
- Innovation-based regulation:** Promotes tech progress while balancing risk and ethical principles.

Policy-making and recommendation

Actionable policy recommendations are essential to promote digital well-being, build trust, and mitigate MDM risks across the digital lifecycle.

- Transparency standards:** mandate genAI developers and users to disclose data sources, usage, and environmental impacts to combat MDM.
- Risk analysis and sequential deployment:** ensure controlled genAI evaluation, balancing public scrutiny with misuse risks for responsible innovation.
- Digital content regulation:** Expand digital content regulation to LLMs and genAI by mandating notice-and-action operating mechanisms, trusted flaggers, and AI-generated content labeling.
- Anti-bias and data for training:** Mandate anti-bias measures in AI training data, ensure representativeness, fairness, and early intervention to mitigate MDM risks.

- Tech-driven vs. global regulation:** Adopt technology-neutral regulations over tech-specific laws to ensure adaptability and global coherence in regulating genAI and LLMs.
- Law for AI-based MDM:** Evolve legal frameworks to mandate AI-generated content labeling, detection mechanisms, and scrutiny to combat MDM risks effectively.
- Building public trust:** enforcing transparency in content moderation, labeling AI-enhanced MDM, and incentivizing digital platforms to prioritize fact-checking and content verification.

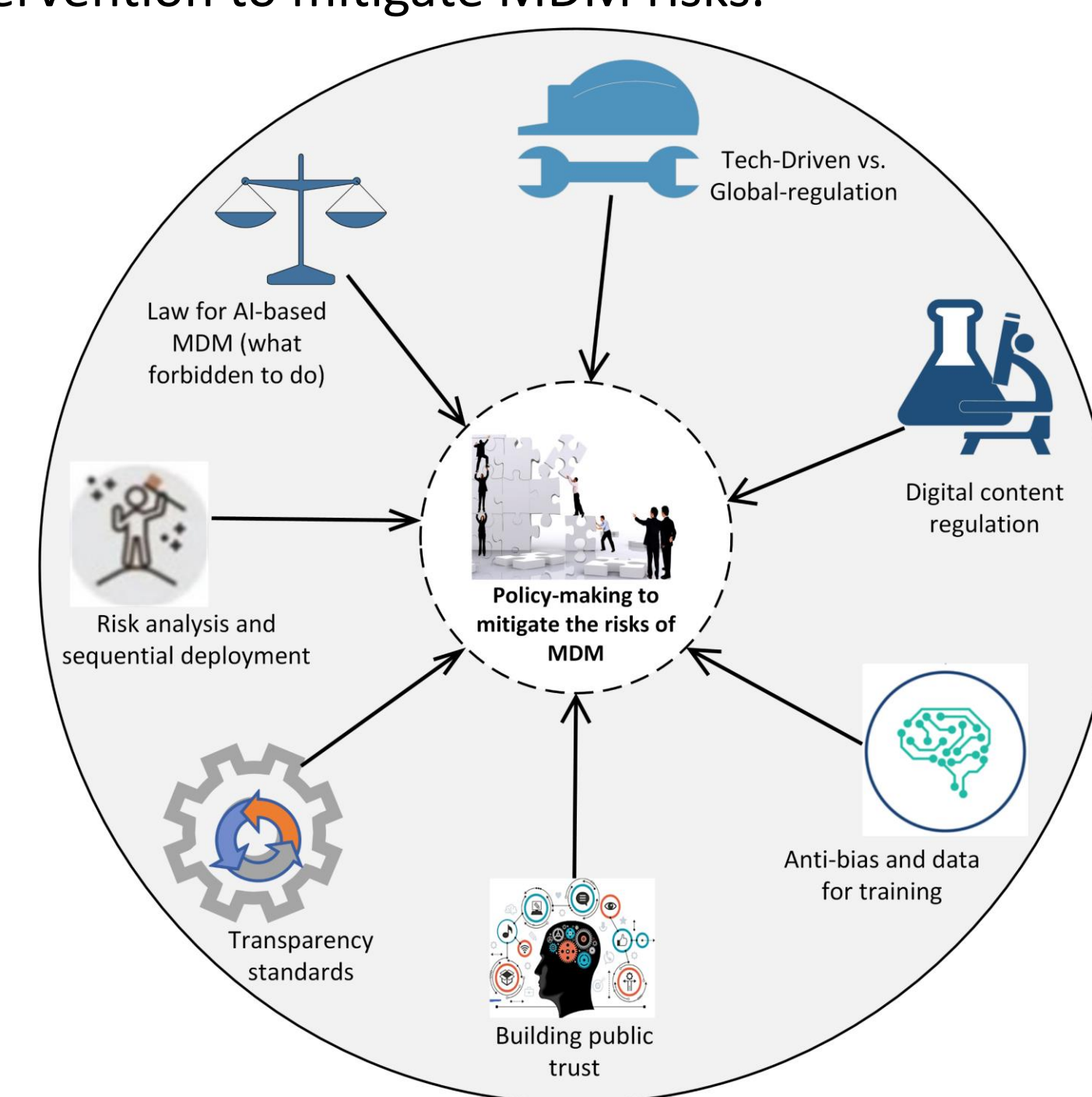


Fig. 4 Policy-making to enhance digital resilience

Discussion and conclusion

- The study highlights critical gaps in global regulatory frameworks and provides actionable policy recommendations to enhance digital resilience, balance invocation, and combat MDM risks in the genAI era.
- By bridging regulatory gaps and distinguishing synthetic content detection from factual inaccuracies, the study provides a foundation for adaptive governance and public trust in the genAI era.

Key Takeaways

- Global regulatory gaps create inconsistencies in policy enforcement.
- Digital resilience should balance innovation, accountability, and consumer protection.
- Cross-border regulatory coherence is critical for mitigating AI-based MDM.
- Actionable policies should integrate AI tools and behavioral science for effectiveness.
- Regulations should align with societal and technological diversity for impact.
- Future governance frameworks should integrate transparency, fairness, and accountability.
- Trust in AI systems depends on regulatory clarity and consumer awareness.
- This study provides a foundation for policymakers to enhance global digital resilience.