International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Development of AI from Search to Companion: A Literature Review

Jerry S Justianto

Abstract

The main problem of AI evolution (from Search Engine, Recommendation Systems, Generative AI, to AI Companion) is the existence of a consistent trade-off pattern in each phase of development: Potential Filter Bubble (limiting perspective), and triggering Polarization or Addiction (especially in the context of recommendation systems). This Study is important. AI has shifted from a functional tool (search) to a digital companion that prioritizes anthropomorphic (human-like) interactions. Artificial Intelligence (AI) has transformed from a PageRank-based search engine to a digital companion system that prioritizes anthropomorphic interactions. This study examines the evolution of AI through four phases: search engines, recommendation systems, generative AI, and AI companions, by analyzing the benefits, risks, and socio-technical implications at each stage. Using a systematic literature review of 45 publications (2014-2025), this study identifies a consistent trade-off pattern: increased personalization and system autonomy accompanied by the risks of filter bubbles, hallucinations, and artificial intimacy. The theoretical framework of the Media Equation and three-factor anthropomorphism explains why users tend to treat AI like humans, raising challenges of overtrust and cognitive manipulation. In this paper, findings highlight the need for ethical governance based on humans-in-the-loop, algorithmic transparency, and AI literacy for responsible adoption. Increasing personalization and system autonomy are always accompanied by increased risks of filter bubbles, hallucinations, and artificial intimacy.

Keywords:
Systematic Literature Review, Artificial Intelligence, Evolution, Companion

1. Introduction

The rapid development of Artificial Intelligence (AI) over the past two decades represents a fundamental transformation in how humans' access, process, and interact with information systems. Early AI applications primarily function as information retrieval tools, where search engines automate relevance ranking and knowledge discovery at unprecedented scales. Foundational studies on hyperlink-based algorithms, such as PageRank, establish the technical and conceptual basis for modern AI-driven information access by formalizing authority, relevance, and trust within large-scale networks [1], [2]. This phase positions AI as a passive yet powerful intermediary, optimizing search efficiency while reshaping information visibility and access structures across the digital ecosystem.

As AI systems evolve, the literature documents a clear transition from static search toward adaptive recommendation systems. These systems leverage machine learning and deep neural networks to model user preferences, behaviors, and engagement patterns in real time. Research demonstrates that recommendation algorithms significantly increase user retention and consumption by personalizing content streams across platforms such as social media, e-commerce, and video services [3], [4]. However,

Corresponding Author: Jerry S Justianto, Binus University (jerry.justiantobinus@binus.ac.id)

the SLR literature consistently highlights structural issues arising from algorithmic curation, including the reinforcement of confirmation bias, opacity of decision-making, and concentration of informational power within platform providers [5]. These concerns motivate a deeper inquiry into how AI-mediated personalization reshapes public discourse.

The next major phase in AI development emerges with generative models capable of producing human-like text, images, code, and multimodal outputs. Advances in transformer architectures and large language models fundamentally shift AI from content filtering to content creation [6], [7]. Empirical studies report substantial productivity gains in knowledge-intensive tasks, with controlled experiments showing performance improvements of up to 40% in writing, coding, and problem-solving activities [8]. Despite these benefits, the literature emphasizes critical limitations, including hallucination, factual inconsistency, and user overreliance, which challenge traditional notions of authorship, accountability, and epistemic trust [9], [10].

More recently, AI systems have evolved into companion-oriented agents that emphasize continuous interaction, emotional responsiveness, and social presence. Companion AI integrates natural language processing, affective computing, and behavioral modeling to simulate empathy, companionship, and support [11], [12]. Studies indicate that users increasingly attribute human-like traits, emotions, and intentions to these systems, reinforcing anthropomorphic engagement patterns. While this evolution expands accessibility to emotional support and personalized assistance, the literature raises ethical and psychological concerns related to artificial intimacy, dependency formation, and blurred boundaries between authentic and synthetic relationships [13].

From a socio-technical perspective, scholars argue that each evolutionary phase of AI introduces a trade-off between functional efficiency and systemic risk. Search engines democratize access to information but contribute to filter bubbles and selective exposure [14], [15]. Recommendation systems optimize engagement but correlate with political polarization and misinformation amplification [16], [17]. Generative AI enhances creativity and productivity while introducing epistemic instability, and companion AI strengthens user engagement while increasing vulnerability to emotional manipulation [12], [18]. These recurring trade-offs underline the need for a holistic analytical framework rather than isolated technological evaluations.

Historical and policy-oriented literature further underscores the urgency of examining AI evolution comprehensively. Harari frames AI-driven information systems as large-scale persuasion infrastructures capable of shaping attention, emotions, and behavior without explicit coercion [19]. Governance studies echo this concern, emphasizing that opaque AI systems may erode autonomy, democratic deliberation, and cognitive agency if left unchecked [20], [21]. These arguments position AI not merely as a technical artifact but as a structural force embedded within political, economic, and cultural systems.

Empirical evidence on social and psychological impacts presents a nuanced and often contested picture. While some studies confirm the presence of echo chambers and algorithmic bias, others suggest that exposure to diverse content does not uniformly reduce polarization and may, in some contexts, intensify it [15], [16], [22]. Public health literature further links prolonged engagement with algorithmic platforms to mental well-being risks, particularly among adolescents and vulnerable populations [23]. This complexity signals that the effects of AI systems are context-dependent, reinforcing the need for systematic synthesis rather than fragmented empirical conclusions.

This Systematic Literature Review addresses these challenges by offering a comprehensive, evolutionary analysis of AI development from search engines to companion AI. The primary contribution of this study lies in integrating technical, psychological, and socio-technical perspectives into a unified framework that maps benefits, risks, and interaction mechanisms across four distinct AI phases. By synthesizing evidence across disciplines, this review aims to identify recurring patterns, critical inflection points, and design implications that inform responsible AI development, organizational adoption, and future research directions [24], [25].

2. Research methods

This study applies a Systematic Literature Review (SLR) methodology guided by the PRISMA protocol to ensure transparency, rigor, and reproducibility in the review process [20]. We adapt the standard PRISMA framework to accommodate the interdisciplinary nature of information technology and artificial intelligence research, where technical, social, and behavioral studies frequently intersect [21]. We conduct the search and selection process in October 2025 using major academic databases, including Scopus, Web of Science, ACM Digital Library, IEEE Xplore, and Google Scholar, to ensure comprehensive coverage of high-quality and reputable journal articles as well as peer-reviewed conference proceedings.

We design the search strategy using structured Boolean combinations to capture the evolutionary trajectory of AI technologies. The primary keywords include ("artificial intelligence" OR "AI" OR "machine learning") AND ("evolution" OR "transformation" OR "development") AND ("search engine" OR "recommendation system" OR "generative AI" OR "large language model" OR "AI companion" OR "conversational AI"). To address the socio-technical and psychological dimensions of AI development, we further incorporate complementary keywords such as "anthropomorphism," "parasocial relationship," "productivity," "mental health," "polarization," and "filter bubble." This multi-layered keyword strategy enables the identification of studies that examine both technological advancements and their broader societal implications.

To strengthen the completeness and reliability of the literature retrieval process, we supplement database searches with manual exploration through Google Scholar and AI-assisted academic discovery tools. We utilize Perplexity Pro to support journal identification by prompting searches such as “provide academic journals from Scopus or SINTA related to” the predefined keywords, ensuring alignment with reputable indexing standards. This combined approach minimizes the risk of missing relevant studies, reduces selection bias, and enhances the robustness of the final literature corpus used for synthesis and analysis.

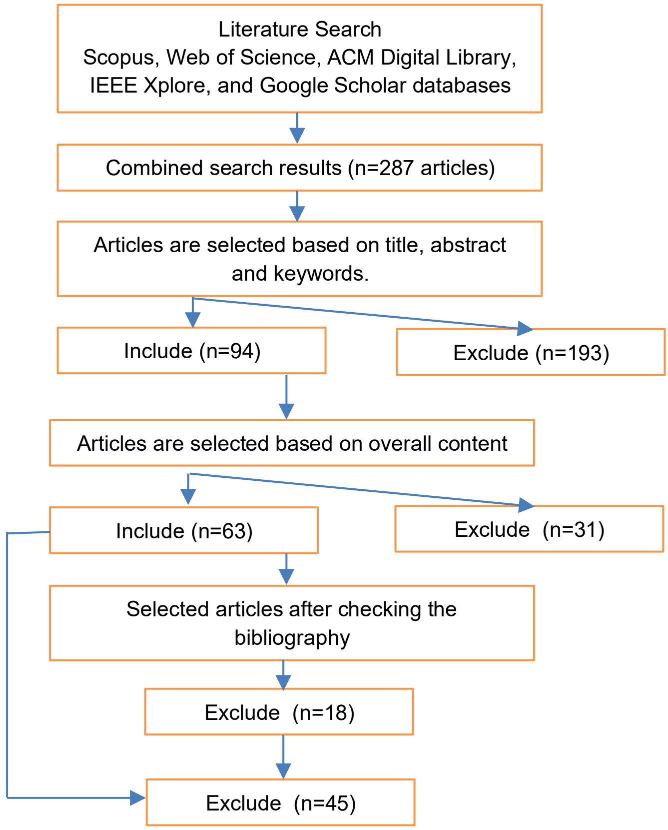

Fig. 1. PRISMA flowchart

According to the 287 identified articles, 94 passed abstract screening, and 45 met the final inclusion criteria after full-text assessment. Data were extracted using a thematic framework encompassing the following dimensions: (1) core roles/functions of technology; (2) measurable benefits; (3) risks and drawbacks; (4) psychological mechanisms; and (5) organizational implications. Content analysis was conducted using deductive-inductive coding to identify patterns across evolutionary phases [22].

During the article selection process, a scoring system was applied. This scoring was used to determine inclusion and exclusion criteria.

Table 1: Inclusion and Exclusion Criteria.

Inclusion	**Exclusions**
Publications for the 2014-2025 period	Non-peer-reviewed articles without a clear institutional affiliation
Scopus/WoS-indexed journal articles, tier-1 conference proceedings (ACM CHI, RecSys, FAccT, AAAI), academic books, and working papers from reputable institutions (NBER, HBS, MIT)	Purely algorithmic technical focus without discussion of social/organizational implications
Direct relevance to one or more phases of AI evolution and/or impact on users/organizations	Preprints without peer review
English language	Duplicate publications.
Full-text available	

Table 2. Assessment Criteria

No	Questions
1	Clear research objectives
2	Research design objectives
3	Clear discussion of variables
4	Clear research steps
5	Clear data collection methods
6	Clear measures of reliability and validity
7	Clear discussion of statistics
8	Clear discussion and results
9	Impact on your research

Selected journals will be assessed based on the nine items above. Each question on the assessment checklist (as shown in Table 4) is scored on a three-point scale (Yes = 1 point, Some = 0.5 points, No = 0 points). Selected studies will be scored from 0 to 9. The higher the percentage, the greater the journal's value to researchers. Table 3 presents the scoring criteria based on nine identified criteria for assessing the quality of the 45 selected studies.

Table 3. Quality Assessment Results

No	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Percentage
1	1	1	1	1	0,5	1	1	0	1	83%
2	1	1	1	1	1	1	0,5	1	1	94%
3	1	1	1	1	1	1	1	1	0,5	94%
4	1	1	1	1	0,5	1	1	0,5	0,5	83%
5	1	1	1	1	0,5	1	1	0	1	83%
6	1	1	1	1	1	0,5	1	1	0,5	89%
7	1	1	1	1	1	1	1	1	1	100%
8	1	1	1	1	1	1	0,5	0,5	1	89%
9	1	1	1	1	1	1	1	1	1	100%
10	1	1	1	1	1	1	1	0,5	0,5	89%
11	1	1	1	1	1	0,5	1	0,5	1	89%
12	1	1	1	1	1	1	1	1	1	100%
13	1	1	1	1	1	1	1	0,5	1	94%
14	1	1	1	1	1	1	1	0,5	1	94%
15	1	1	1	1	1	1	1	1	1	100%
16	1	1	1	1	1	1	1	1	0,5	94%
17	1	1	1	1	1	1	0,5	1	1	94%
18	1	1	1	1	1	1	1	0,5	0,5	89%
19	1	1	1	1	0,5	1	1	0,5	0,5	83%
20	1	1	1	1	1	0,5	1	1	1	94%
21	1	1	1	1	1	1	1	1	1	100%
22	1	1	1	1	0,5	1	1	0,5	1	89%
23	1	1	1	1	1	0,5	1	1	1	94%
24	1	1	1	1	1	1	1	1	1	100%
25	1	1	1	1	1	1	1	0,5	0,5	89%
26	1	1	1	1	1	0,5	1	0,5	1	89%
27	1	1	1	1	1	1	1	1	1	100%
28	1	1	1	1	1	1	1	1	1	100%
29	1	1	1	1	0,5	1	1	0,5	0,5	83%
30	1	1	1	1	1	1	1	0,5	1	94%
31	1	1	1	1	0,5	1	1	0,5	0,5	83%
32	1	1	1	1	1	0,5	0,5	0,5	1	83%
33	1	1	1	1	1	1	1	0,5	0,5	89%
34	1	1	1	1	1	1	1	0,5	1	94%
35	1	1	1	1	1	1	1	0,5	1	94%
36	1	1	1	1	1	1	1	1	1	100%

37	1	1	1	1	0,5	1	1	0	1	83%
38	1	1	1	1	1	1	0,5	1	1	94%
39	1	1	1	1	1	1	1	1	0,5	94%
40	1	1	1	1	0,5	1	1	0,5	1	89%
41	1	1	1	1	1	1	1	0,5	0,5	89%
42	1	1	1	1	1	0,5	1	0,5	1	89%
43	1	1	1	1	1	1	1	0,5	0,5	89%
44	1	1	1	1	1	0,5	1	0,5	1	89%
45	1	1	1	1	1	1	1	0,5	0,5	89%

A quality assessment of the 45 selected studies found that all scored above 60%, thus passing the quality assessment. Twenty percent of the studies scored 100%, 28.8% scored 94%, and 51.2% scored 80-89%. The following section presents an in-depth analysis of the data collected from the 45 selected studies.

3. Results and Discussion

Table 4. Previous research

AI Evolutionary Phases	**Study Examples & Ref.**	**Core Roles/Functions**	**Key Measurable Benefits**	**Key Risks & Weaknesses**
Search Engines	Brin & Page [2], Lewandowski [23], White & Roth [24]	Information Access & Retrieval (PageRank)	Search Efficiency, Democratization of Knowledge.	Filter Bubble (contextual), Limitations of Exploratory Search.
Recommendation Systems	Covington et al. [3], Zhang et al. [4], Gomez-Uribe & Hunt [28]	Proactive Content Discovery (Push, Deep Learning)	Increased Engagement (20-40%), Revenue Impact (35-75% of consumption).	Political Polarization (Bail et al. [10]), Backfire Effect (Nyhan & Reifler [30]), Addiction.
Generative AI	Bommasani [5], Noy & Zhang [12], Dell'Acqua et al. [14]	Multimodal Content Creation (Generative)	Increased Productivity (14-40%), Skill-Leveling (Brynjolfsson et al. [13]).	Hallucinations (3-27% of output), Overreliance, Jagged Frontier [14].
AI Companion	Skjuve et al. [6], Ta et al. [7]	Emotional Support, Anthropomorphic Interaction	24/7 Availability, Non-judgmental, Potential for Behavior Change Interventions.	Artificial Intimacy, Parasocial Relationships, Reduction of Real Social Skills.

Phase 1: Search Engines – From Pull to Universal Access

Early search engines, particularly Google with its PageRank algorithm, fundamentally transform how users retrieve and evaluate information by ranking web pages based on hyperlink structures and query relevance [2]. This technological breakthrough shifts information access from scarce and institutionally mediated resources to an open, globally accessible system. As a result, users retrieve vast amounts of information within seconds, enabling rapid decision-making and supporting a wide range of academic, professional, and everyday activities [23]. The literature consistently highlights that this phase of AI-driven information systems establishes the foundation for later intelligent systems by formalizing relevance, authority, and ranking as core computational problems.

Empirical studies demonstrate that search engines dramatically increase information retrieval efficiency compared to pre-digital methods, such as manual library searches or static directories [24]. Researchers report that search technologies support self-directed learning by allowing users to explore topics independently, refine queries iteratively, and access diverse sources at minimal cost [25]. In organizational contexts, search engines reduce the transaction costs of information access to near zero, which significantly enhances knowledge worker productivity and lowers barriers to innovation. These measurable benefits explain why search-based AI systems quickly become embedded in daily life and institutional practices.

Despite these advantages, the personalization of search results introduces concerns regarding informational narrowing and selective exposure. Scholars argue that algorithmic personalization may lead to filter bubbles, where users predominantly encounter information that reinforces existing beliefs and preferences, potentially limiting exposure to alternative viewpoints and critical discourse [8]. This issue raises broader questions about epistemic diversity, democratic deliberation, and the role of algorithmic systems in shaping public knowledge.

However, empirical findings present a more nuanced picture of these risks. Studies by Bruns [9] and Zuiderveen Borgesius et al. [18] suggest that filter bubbles are generally weaker and more context-dependent than commonly assumed, with significant variation driven by users' digital literacy, media habits, and intentional information-seeking behavior. Haim et al. [26] further demonstrate that diversity in news consumption depends more strongly on individual preferences than on algorithmic ranking alone. These findings indicate that while early search engines introduced structural risks related to personalization, user agency, and contextual factors play a substantial role in shaping actual information exposure.

Phase 2: Recommendation Systems – From Pull to Push

Recommendation systems fundamentally shift the information interaction paradigm from pull-based models, where users actively search for content, to push-based models, where systems proactively suggest items based on inferred preferences and behavior. Platforms such as YouTube operationalize this shift through a two-stage architecture that combines deep candidate generation using collaborative filtering neural networks with a subsequent deep ranking phase that evaluates hundreds of contextual and behavioral features [3]. This architecture becomes an industry blueprint for large-scale content discovery systems and demonstrates how deep learning enables personalization at unprecedented scale. Empirical studies consistently show that deep learning–based recommendation systems outperform traditional rule-based or heuristic approaches.

Quantitative evidence highlights the substantial performance gains achieved through recommendation systems. Research reports that personalized recommendations increase user engagement by approximately 20–40% compared to conventional non-personalized methods [4, 27]. These systems enable efficient content discovery by allowing users to find relevant items without the need to formulate explicit queries, thereby reducing cognitive effort and interaction friction [28]. As a result, recommendation-driven platforms achieve faster time-to-value and higher user satisfaction scores, which directly translate into improved retention and long-term platform loyalty [45].

From an economic perspective, recommendation systems exert a significant impact on platform revenue and content consumption patterns. Industry analyses indicate that 35–75% of Netflix's total content consumption originates from algorithmic recommendations rather than direct search or manual browsing [28]. This dominance illustrates how recommendation engines shape user attention and influence market dynamics by amplifying certain content while marginalizing others. Consequently, recommendation systems become central not only to user experience design but also to business sustainability and competitive advantage.

However, the social and psychological implications of recommendation systems remain complex and contested. Bail et al. [10] demonstrate through controlled experiments that exposure to opposing political views on Twitter can increase polarization, particularly among conservative users, due to a backfire effect in which counter-attitudinal information reinforces prior beliefs [30]. In contrast, Guess et al. [31] apply counterfactual bot analysis and find that algorithmic recommendations contribute less to polarization than widely assumed, with user-driven selective exposure playing a more dominant role. In the domain of mental well-being, a systematic review by Shannon et al. [32] covering 87 studies identifies a modest association between intensive social media use and depression or anxiety ($r = 0.15–0.25$), with substantial heterogeneity based on usage patterns and content types. Complementary findings by Conte et al. [34] on TikTok usage among 17,336 adolescents reveal elevated risks of problematic use, body image concerns, and reduced life satisfaction, while emphasizing the need for longitudinal causal research to fully understand these dynamics.

Phase 3: Generative AI – Accelerating Productivity and the Jagged Frontier

The emergence of Large Language Models (LLMs) such as GPT-3/4, Claude, and Gemini represents a qualitative leap in the evolution of artificial intelligence, fundamentally redefining how machines process, generate, and interact with information. Unlike earlier AI systems that primarily rely on retrieval, ranking, or pattern matching over existing content, LLMs operate as generative models that synthesize new outputs based on probabilistic language representations learned from massive-scale datasets [34, 35]. These models demonstrate an advanced capacity to generate coherent, context-aware, and semantically rich text, as well as executable code, structured reasoning steps, and increasingly multimodal content that integrates language with images, audio, and video. The literature emphasizes that this generative capability shifts AI from a supportive information tool into an active cognitive collaborator, capable of participating in complex tasks such as problem-solving, creative writing, programming assistance, and decision support. As a result, LLMs blur the boundary between information retrieval and knowledge creation, raising both unprecedented opportunities for productivity and creativity and new challenges related to reliability, accountability, and human–AI interaction.

Measurable Productivity Benefits: Experimental evidence consistently demonstrates significant productivity gains. Noy & Zhang [12], in a randomized controlled trial of professional writing tasks, found a 40% decrease in completion time and an 18% increase in output quality (assessed by an independent evaluator). Brynjolfsson et al. [13] analyzed operational data from 5,179 customer support agents before and after adopting a GenAI assistant and found an average 14% increase in productivity, with the largest effect (35%) among less experienced workers. This indicates a diffusion of knowledge from high-performers to low-performers through AI mediation (a skill-leveling effect). A further field experiment study by Peng et al. [36] on business consultants found a 12.2% improvement in quality and 25.1% in speed for tasks in the capability frontier model.

The risk of jagged frontiers and hallucinations: The concept of a "jagged technological frontier" [14] explains that LLMs are highly capable within their training domain, but their performance drops dramatically or they produce false but convincing output (hallucinations) outside that domain. In an experiment with 758 consultants using GPT-4, those given unguided access to the AI for tasks outside the frontier experienced a 19%

performance decline compared to controls, due to overconfidence in the erroneous output. Ji et al. [37], in a comprehensive survey of hallucinations in LLMs, identified that 3-27% of output contains factually incorrect information, depending on the domain and model. This requires sophisticated human-in-the-loop verification and prompting strategies [38].

Phase 4: AI Companion – Simulated Closeness Between Support and Artificial Intimacy

AI companions such as Replika, Character.AI, and Pi combine generative language capabilities with persona design that optimizes empathy cues, active listening responses, and self-disclosure prompts to foster interpersonal bonding [6,7]. A qualitative study of 15 long-term Replika users by Skjuve et al. [6] found that the majority developed a parasocial relationship with the AI—a sense of companionship, emotional attachment, and disclosure of intimacy comparable to human relationships, despite cognitively recognizing that the partner was nonhuman.

Potential mental health benefits: A systematic review by Abd-Alrazaq et al. [40] of 11 RCTs on mental health chatbots found small-to-moderate effects for reducing symptoms of depression and anxiety in a mild-to-moderate clinical population. Laranjo et al. [40], in a meta-analysis of 17 studies, found chatbots to be effective for behavior change interventions (treatment adherence, physical activity) with an effect size of $d = 0.38$. However, the quality of evidence is heterogeneous, and no studies have shown that chatbots can replace human professionals for serious conditions.

Risks of artificial intimacy and manipulation: Parasocial relationship theory [41, 42] suggests that humans can develop unilateral emotional bonds with media personas. When applied to responsive and adaptive AI companions, the greatest risk is illusory understanding—users feeling deeply understood when in fact the AI is simply pattern-matching without genuine empathy or consciousness [43]. This has the potential to make misleading or manipulative information appear persuasive due to its warm and personalized presentation. Brown et al. [35] found that 64% of AI companion users reported following AI advice for important decisions without external verification—an indication of dangerous overtrust.

From a mental health perspective, the Office of the U.S. Surgeon General (2023) warns that while AI companions can provide transactional support, excessive reliance can reduce investment in real human relationships and social skills development, particularly in adolescent populations and individuals with attachment issues. Additional risks include data privacy (intensive disclosure of sensitive information), manipulation via reinforcement schedules, and difficulty disengaging from relationships in which one is deeply emotionally invested.

Comparative analysis identified relatively consistent trade-off patterns throughout AI evolution. Each phase brings tangible functional improvements: information access (Search), content discovery (Recommendation), productivity (Generative AI), and emotional support (Companion). However, these benefits come at the cost of increased risks to cognitive autonomy, psychological well-being, and individual agency.

Table 5. Main trade-off patterns:

Faze	Core Functions	Primary Benefits	Primary Risks	Mitigation Strategy
Search Engines	Access information	Efficiency, democratization of knowledge	Filter bubble (contextual)	Diversity nudges, transparency tools
Recommendations	Content discovery	Engagement, personalization	Polarization (conditional), addiction	Explanability, dwell time limits

Generative AI	Content creation	Productivity (+14-40%), skill leveling	Hallucination, overreliance	Human-in-the-loop, verification protocols
AI Companion	Emotional support	24/7 availability, non-judgmental	Artificial intimacy, overtrust	Boundary disclosure, escalation paths

Harari's framework [15] provides a normative lens: as information networks become "human-faced" persuasion mediums, the risk of manipulation increases not because AI has physical agency, but because anthropomorphic interfaces slip into human cognitive and affective gaps. Designs that optimize engagement without safeguards can create dependencies that erode individual autonomy.

4. Conclusion

This systematic literature review synthesizes evidence across four major phases of AI evolution, including search engines, recommendation systems, generative AI, and AI companions. It can demonstrate a consistent pattern of socio-technical trade-offs accompanying technological progress. The findings confirm that early search engines fundamentally democratize information access and dramatically increase retrieval efficiency, forming the computational foundation for later intelligent systems. While concerns about filter bubbles and selective exposure persist, empirical studies show that these risks are highly context-dependent and strongly mediated by user agency, digital literacy, and information-seeking behavior. Thus, the first phase illustrates that AI-driven efficiency gains do not deterministically undermine epistemic diversity, but instead interact with human choices and institutional contexts.

The review further reveals that recommendation systems and generative AI mark a qualitative escalation in both benefits and risks. Recommendation systems successfully shift interaction from pull to push models, producing substantial gains in engagement, satisfaction, and economic value, yet they also amplify concerns related to polarization, attention manipulation, and mental well-being. Generative AI, particularly large language models, delivers unprecedented productivity improvements—ranging from 14% to 40% depending on task and expertise level. It can simultaneously introduce the “jagged frontier” problem, where high apparent competence coexists with non-trivial hallucination rates and performance degradation outside core domains. These findings underscore that productivity gains are inseparable from new forms of cognitive risk, especially overreliance and misplaced trust, necessitating robust human-in-the-loop mechanisms and verification protocols.

Therefore, the emergence of AI companions represents the most profound shift in human–AI interaction, moving beyond functional assistance toward emotional and relational engagement. The literature indicates modest mental health and behavioral benefits in specific contexts, but also highlights significant risks related to artificial intimacy, over trust, and reduced human social investment. Across all phases, the review identifies a recurring structural pattern: each technological advance increases convenience, personalization, and support while simultaneously challenging cognitive autonomy, psychological well-being, and individual agency. In line with Harari’s normative framework, this study concludes that future AI development must prioritize adaptive governance, transparent design, and boundary-aware interaction models to ensure that increasingly human-like AI systems augment human capabilities without undermining autonomy or social resilience.

References

[1] Norvig, Peter, and S. Russell. "Artificial Intelligence: A Modern Approach, 4th US ed." (2021).

[2] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." *Computer networks and ISDN systems* 30, no. 1-7 (1998): 107-117.

[3] Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." In Proceedings of the 10th ACM conference on recommender systems, pp. 191-198. 2016.

[4] Zhang, Shuai, Lina Yao, Aixin Sun, and Yi Tay. "Deep learning based recommender system: A survey and new perspectives." ACM computing surveys (CSUR) 52, no. 1 (2019): 1-38.

[5] Bommasani, Rishi. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

[6] Skjuve, Marita, Ida Maria Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg. "Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction." Human technology 15, no. 1 (2019): 30-54.

[7] Ta, Vivian, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. "User experiences of social support from companion chatbots in everyday contexts: thematic analysis." Journal of medical Internet research 22, no. 3 (2020): e16235.

[8] Pariser, Eli. The filter bubble: What the Internet is hiding from you. penguin UK, 2011.

[9] Bruns, A. Filter bubble. Internet Policy Review, 8(4), (2019). 1-14.

[10] Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences* 115, no. 37 (2018): 9216-9221.

[11] Geschke, Daniel, Jan Lorenz, and Peter Holtz. "The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers." British Journal of Social Psychology 58, no. 1 (2019): 129-149.

[12] Noy, Shakked, and Whitney Zhang. "Experimental evidence on the productivity effects of generative artificial intelligence." Science 381, no. 6654 (2023): 187-192.

[13] Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "Generative AI at work." The Quarterly Journal of Economics 140, no. 2 (2025): 889-942.

[14] Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." Harvard Business School Technology & Operations Mgt. Unit Working Paper 24-013 (2023).

[15] Harari, Yuval Noah. Nexus: A brief history of information networks from the Stone Age to AI. Random House, 2024.

[16] Hagendorff, Thilo. "The ethics of AI ethics: An evaluation of guidelines." Minds and machines 30, no. 1 (2020): 99-120.

[17] Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." Nature machine intelligence 1, no. 9 (2019): 389-399.

[18] Zuiderveen Borgesius, Frederik J., Damian Trilling, Judith Möller, Balázs Bodó, Claes H. De Vreese, and Natali Helberger. "Should we worry about filter bubbles?." *Internet policy review* 5, no. 1 (2016).

[19] Twenge, Jean M., Jonathan Haidt, Andrew B. Blake, Cooper McAllister, Hannah Lemon, and Astrid Le Roy. "Worldwide increases in adolescent loneliness." Journal of adolescence 93 (2021): 257-269.

[20] Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." bmj 372 (2021).

[21] Kitchenham, Barbara, and Stuart Charters. "Guidelines for performing systematic literature reviews in software engineering." (2007).

[22] Braun, Virginia, and Victoria Clarke. "Using thematic analysis in psychology." *Qualitative research in psychology* 3, no. 2 (2006): 77-101.

[23] Lewandowski, Dirk. "The retrieval effectiveness of web search engines: considering results descriptions." Journal of documentation 64, no. 6 (2008): 915-937.
[24] White, Ryen W., and Resa A. Roth. Exploratory search: Beyond the query-response paradigm. No. 3. Morgan & Claypool Publishers, 2009.
[25] Marchionini, Gary. "Exploratory search: from finding to understanding." Communications of the ACM 49, no. 4 (2006): 41-46.
[26] Haim, Mario, Andreas Graefe, and Hans-Bernd Brosius. "Burst of the filter bubble? Effects of personalization on the diversity of Google News." Digital journalism 6, no. 3 (2018): 330-343.
[27] He, Xiangnan, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering." In Proceedings of the 26th international conference on world wide web, pp. 173-182. 2017.
[28] Gomez-Uribe, Carlos A., and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation." ACM Transactions on Management Information Systems (TMIS) 6, no. 4 (2015): 1-19.
[29] Lu, Jie, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. "Recommender system application developments: a survey." Decision support systems 74 (2015): 12-32.
[30] Nyhan, Brendan, and Jason Reifler. "When corrections fail: The persistence of political misperceptions." Political Behavior 32, no. 2 (2010): 303-330.
[31] Guess, Andrew M., Pablo Barberá, Simon Munzert, and JungHwan Yang. "The consequences of online partisan media." Proceedings of the National Academy of Sciences 118, no. 14 (2021): e2013464118.
[32] Shannon, Holly, Katie Bush, Paul J. Villeneuve, Kim GC Hellemans, and Synthia Guimond. "Problematic social media use in adolescents and young adults: systematic review and meta-analysis." JMIR mental health 9, no. 4 (2022): e33450.
[33] Conte, Giulia, Giorgia Di Iorio, Dario Esposito, Sara Romano, Fabiola Panvino, Susanna Maggi, Benedetta Altomonte, Maria Pia Casini, Mauro Ferrara, and Arianna Terrinoni. "Scrolling through adolescence: a systematic review of the impact of TikTok on adolescent mental health." European Child & Adolescent Psychiatry 34, no. 5 (2025): 1511-1527.
[34] Bommasani, Rishi, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, and Percy Liang. "The foundation model transparency index." *arXiv preprint arXiv:2310.12941* (2023).
[35] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
[36] Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. "The impact of ai on developer productivity: Evidence from github copilot." arXiv preprint arXiv:2302.06590 (2023).
[37] Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." ACM computing surveys 55, no. 12 (2023): 1-38.
[38] White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. "A prompt pattern catalog to enhance prompt engineering with chatgpt." arXiv preprint arXiv:2302.11382 (2023).
[39] Abd-Alrazaq, Alaa Ali, Asma Rababeh, Mohannad Alajlani, Bridgette M. Bewick, and Mowafa Househ. "Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis." Journal of medical Internet research 22, no. 7 (2020): e16021.
[40] Laranjo, Liliana, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian et al. "Conversational agents in healthcare: a systematic review." Journal of the American Medical Informatics Association 25, no. 9 (2018): 1248-1258.
[41] Horton, Donald, and R. Richard Wohl. "Mass communication and para-social interaction: Observations on intimacy at a distance." psychiatry 19, no. 3 (1956): 215-229.

[42] Dibble, Jayson L., Tilo Hartmann, and Sarah F. Rosaen. "Parasocial interaction and parasocial relationship: Conceptual clarification and a critical assessment of measures." Human communication research 42, no. 1 (2016): 21-44.
[43] Maeda, Takuya, and Anabel Quan-Haase. "When human-AI interactions become parasocial: Agency and anthropomorphism in affective design." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp. 1068-1077. 2024.
[44] Resnick, Paul, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. "Bursting your (filter) bubble: strategies for promoting diverse exposure." In Proceedings of the 2013 conference on Computer supported cooperative work companion, pp. 95-100. 2013.
[45] Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi. "Recommender systems and their ethical challenges." Ai & Society 35, no. 4 (2020): 957-967.
[46] Lau, Josephine, Benjamin Zimmerman, and Florian Schaub. "Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers." Proceedings of the ACM on human-computer interaction 2, no. CSCW (2018): 1-31.
[47] Natale, Simone. Deceitful media: Artificial intelligence and social life after the Turing test. Oxford University Press, 2021.