International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 118-128

NotebookLM as a Source-Grounded AI Research Assistant: A Narrative Review of Cognitive Load Implications and Research Originality Concerns

Fatima M. Salman, Samy S. Abu-Naser

Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine

Abstract: *The proliferation of AI-powered tools in academic settings has prompted urgent questions about their cognitive and epistemic effects on researchers. Among emerging tools, Google's NotebookLM occupies a distinctive position: unlike general-purpose large language models (LLMs) such as ChatGPT or Gemini, it operates exclusively within a user-defined document corpus, grounding every response in cited, user-supplied sources. This architecture raises theoretically significant questions about cognitive load and research originality — two constructs central to the quality and integrity of academic knowledge production. This paper presents a critical narrative review of the literature at the intersection of source-grounded AI, Cognitive Load Theory (CLT), and academic originality, with particular attention to what current evidence suggests — and fails to address, about NotebookLM specifically. Drawing on 45 peer-reviewed studies and technical reports published between 2015 and 2024, we identify four major themes: (1) the cognitive affordances and constraints of source-grounded versus open-ended AI retrieval; (2) the applicability of CLT's load taxonomy to AI-mediated research tasks; (3) empirical evidence on AI's effects on creative and scholarly originality; and (4) the methodological and ethical challenges of studying AI-human research collaboration. We synthesise these themes into a conceptual framework — the Grounded Retrieval–Cognitive Redistribution (GRCR) model — that predicts when and for whom source-grounded AI tools are likely to reduce extraneous cognitive load without displacing germane elaboration. We conclude by identifying critical gaps in the literature and proposing a research agenda for empirical investigation of NotebookLM and similar grounded-retrieval systems in academic contexts.*

Keywords: NotebookLM, Cognitive Load Theory, Research Originality, Source-Grounded AI, Large Language Models, Academic Research, AI-Assisted Knowledge Work.

1. Introduction

The past decade has witnessed an unprecedented acceleration in the integration of artificial intelligence (AI) into academic research workflows. What began with algorithmic recommendation engines embedded in citation managers has evolved into fully conversational, generative systems capable of synthesising literature, formulating hypotheses, and drafting scholarly prose at scale (Bender et al., 2021; Susarla et al., 2023). Within this rapidly expanding ecosystem, one architectural distinction has emerged as theoretically and practically significant: the difference between open-ended large language models (LLMs) that draw on vast parametric world knowledge, and source-grounded systems that restrict their outputs to a user-defined corpus of documents.

Google's NotebookLM, released in 2023 and substantially updated in 2024, exemplifies the latter category. Designed explicitly as a "personal AI research assistant," NotebookLM accepts user-uploaded documents — journal articles, reports, lecture notes, books — and generates all responses by retrieving and synthesising content exclusively from those sources, providing inline citations to specific passages (Google, 2024). This closed-corpus architecture has been positioned by its developers as a safeguard against hallucination and a mechanism for preserving source traceability, both of which are essential properties for scholarly work (Rawte et al., 2023). Yet despite these design commitments, empirical understanding of how NotebookLM affects the researchers who use it — cognitively, creatively, and epistemically — remains critically underdeveloped.

Two constructs are particularly central to this understanding. The first is cognitive load, conceptualised within Sweller's (1988) Cognitive Load Theory (CLT) as the demands placed on working memory during complex tasks. If source-grounded retrieval reduces the cognitive cost of locating, verifying, and cross-referencing sources, it may free researchers to invest more working memory in the higher-order activities — synthesis, critique, theorisation — that generate scholarly value. The second is research originality, encompassing the novelty, conceptual depth, and critical independence of academic outputs. A tool that reduces cognitive burden but simultaneously constrains researchers to pre-selected sources may inadvertently narrow the epistemic horizon from which original ideas can emerge.

This paper addresses the gap between NotebookLM's theoretical promise and the absence of systematic empirical investigation through a critical narrative review of the interdisciplinary literature spanning AI-assisted knowledge work, cognitive load research, creativity science, and library and information science. Our review is organised around four overarching questions:

- What cognitive mechanisms distinguish source-grounded AI retrieval from both unaided research and open-ended LLM assistance, and what does CLT predict about their effects?
- What does existing empirical evidence reveal about the effects of AI tools on researchers' cognitive load during literature-intensive tasks?
- How do AI tools affect the originality of academic outputs, and under what conditions do they enhance or undermine scholarly creativity?
- What methodological and ethical challenges must future empirical studies address when investigating source-grounded AI tools like NotebookLM?

We contribute to these questions in three ways. First, we synthesise a body of literature that has not previously been brought together around the specific case of source-grounded AI in academic settings. Second, we propose the Grounded Retrieval–Cognitive Redistribution (GRCR) model as a conceptual framework for predicting and explaining NotebookLM's cognitive and creative effects. Third, we identify specific, addressable gaps in the literature and propose a concrete research agenda to guide future empirical work.

The remainder of the paper is structured as follows. Section 2 describes our review methodology. Sections 3 through 6 present the four thematic strands of our review. Section 7 introduces the GRCR model and synthesises the thematic findings. Section 8 addresses limitations of the current evidence base, and Section 9 presents our conclusions and research agenda.

2. Review Methodology

This paper adopts a narrative review approach (Ferrari, 2015), which is appropriate for an emerging and interdisciplinary phenomenon where the goal is critical synthesis and theoretical development rather than meta-analytic aggregation of comparable effect sizes. Narrative reviews are particularly well-suited to topics — such as source-grounded AI in academic research — that cut across multiple disciplines and involve heterogeneous study designs, theoretical frameworks, and outcome measures.

2.1 Search Strategy

We conducted systematic searches across five databases: Web of Science, Scopus, ACM Digital Library, IEEE Xplore, and Google Scholar. Search terms were organised into three conceptual clusters combined with Boolean operators: (1) tool terms: "NotebookLM" OR "source-grounded AI" OR "retrieval-augmented generation" OR "grounded language model"; (2) cognitive terms: "cognitive load" OR "mental effort" OR "working memory" OR "cognitive overhead"; (3) academic/originality terms: "academic research" OR "research originality" OR "scholarly creativity" OR "literature review" OR "knowledge synthesis". Searches were conducted in January 2024 and updated in October 2024.

2.2 Inclusion and Exclusion Criteria

Studies were included if they: (a) were published in English in peer-reviewed journals, conference proceedings, or as preprints on recognised repositories (arXiv, SSRN, OSF); (b) addressed at least one of the following — AI tools in research or knowledge work, cognitive load in technology-mediated tasks, or AI effects on creativity and originality; (c) were published between January 2015 and October 2024. Studies were excluded if they focused exclusively on K-12 educational contexts, automated programming tasks without clear relevance to research cognition, or medical/clinical AI without knowledge-work dimensions.

2.3 Corpus

After screening 312 initial records by title and abstract, 98 were retrieved for full-text review. Of these, 45 were included in the final synthesis: 28 empirical studies (experiments, surveys, case studies), 11 theoretical or conceptual papers, and 6 technical reports or white papers. The thematic distribution of included studies is presented in Table 1.

Table 1

Distribution of Included Literature by Theme and Study Type

Thematic Area	Empirical	Theoretical	Technical	Total
Source-grounded AI & Retrieval-Augmented Generation	5	3	4	**12**

Cognitive Load in AI-Mediated Tasks	9	4	1	**14**
AI Effects on Creativity & Originality	8	2	1	**11**
Methodology & Ethics of AI Research Studies	6	2	0	**8**
TOTAL	**28**	**11**	**6**	**45**

Note. Studies addressing multiple themes were classified by their primary contribution.

2.4 Quality Appraisal

To ensure the rigour of this narrative synthesis, the selected literature underwent a quality appraisal process. Empirical studies were evaluated based on their methodological transparency, sample size adequacy, and the clarity of their findings. Theoretical and technical papers were assessed for their conceptual coherence and relevance to the specific architecture of source-grounded systems. While the review includes preprints to capture the most recent developments in AI research (up to October 2024), priority in synthesis was given to peer-reviewed experimental studies with high ecological validity, thereby enhancing the reliability of the derived GRCR model.

3. Source-Grounded AI: Architecture, Affordances, and Constraints

3.1 From Open-Ended to Grounded Generation

The dominant paradigm in conversational AI since the release of GPT-3 (Brown et al., 2020) has been open-ended generation: a model trained on vast corpora of internet text responds to queries by sampling from its parametric knowledge — representations of information encoded in billions of neural network weights. This paradigm offers extraordinary breadth but introduces a fundamental reliability problem. Because the model's "memory" is distributed across weights rather than stored in retrievable documents, responses can be fluent and confident yet factually erroneous — a phenomenon variously termed hallucination (Maynez et al., 2020), confabulation (Rawte et al., 2023), or sycophantic plausibility (Perez-Ortiz et al., 2023).

Retrieval-Augmented Generation (RAG; Lewis et al., 2020) emerged as an architectural response to this limitation. In RAG systems, a retrieval module first identifies relevant passages from an external corpus, which are then provided to the generative model as explicit context. The generative model's task is thereby constrained: it must formulate a response grounded in the retrieved passages rather than relying solely on parametric knowledge. NotebookLM implements a personalised variant of this architecture in which the retrieval corpus is entirely defined by the user's uploaded documents, making it a closed-personal-corpus RAG system.

This design has several implications that are directly relevant to cognitive and epistemic analysis. First, the grounding constraint transforms the hallucination risk from a global problem (the model may assert anything it was trained on) to a local one (the model may misrepresent or over-extrapolate from the user's documents). Second, inline citations enable post-hoc verification — a researcher can check whether a synthesised claim accurately reflects its cited source — which is not possible with open-ended LLMs unless the researcher independently locates the source. Third, the closed-corpus constraint means the tool cannot introduce information from outside the user's document set, which may be a feature (preventing scope creep or off-topic hallucination) or a limitation (preventing beneficial analogical transfer from adjacent fields) depending on the research task.

3.2 NotebookLM's Interface Design and Researcher Interaction

Beyond its underlying architecture, NotebookLM's user interface embodies specific design choices with cognitive implications. The Notebook Guide feature automatically generates a document overview, key topics, and suggested questions upon source upload, offering a pre-structured cognitive scaffold before the researcher has formulated their own queries. The Audio Overview feature synthesises uploaded sources into a podcast-style conversational summary, offering an auditory alternative to text-based reading. The inline citation panel presents source passages alongside generated responses, enabling rapid triangulation between synthesis and source material.

These features collectively suggest a design philosophy oriented toward reducing the initial cognitive barrier to engaging with a new document corpus — what might be termed the orientation cost of unfamiliar literature. Whether this scaffolding supports or supplants the researcher's own cognitive processing of the source material is an empirical question that, as Section 5 will show, the existing literature has not yet directly addressed for NotebookLM specifically.

3.3 Comparison with General-Purpose LLMs in Research Contexts

Studies comparing RAG-based and open-ended LLM outputs in knowledge-intensive tasks have consistently found RAG to produce more factually accurate and source-consistent outputs (Gao et al., 2023; Shuster et al., 2021). In a controlled evaluation of question answering over scientific literature, Shi et al. (2023) found that RAG systems reduced hallucination rates by 38% compared to parametric-only baselines, with accuracy gains concentrated on questions requiring multi-document synthesis. Notably, however, RAG systems showed smaller accuracy advantages on questions requiring cross-domain analogical reasoning — a finding consistent with the theoretical prediction that closed-corpus grounding may constrain the breadth of conceptual resources available to the reasoner.

For research practitioners, these technical distinctions translate into qualitatively different interaction dynamics. With open-ended LLMs, the researcher's primary verification challenge is factual accuracy — determining whether the model's claims correspond to reality. With source-grounded systems like NotebookLM, the primary challenge shifts to coverage adequacy — determining whether the researcher's uploaded corpus is sufficiently comprehensive to support the synthesis task at hand. This shift in cognitive demand has not been examined empirically in academic user populations.

4. Cognitive Load Theory and AI-Mediated Research Tasks

4.1 CLT Foundations and the Three-Load Taxonomy

Cognitive Load Theory (CLT), developed by Sweller (1988) and elaborated across three decades of empirical and theoretical work (Sweller et al., 2019; Paas et al., 2003), provides the primary theoretical framework for understanding how cognitive resources are allocated during complex tasks. CLT's central premise is that human working memory is severely limited in both capacity (approximately four chunks of novel information; Cowan, 2001) and duration (information decays without active maintenance). Effective performance on complex tasks — including literature review and knowledge synthesis — depends on efficiently managing working memory demands.

CLT distinguishes among three types of cognitive load. Intrinsic load arises from the inherent complexity of the task material — the number of interacting elements that must be processed simultaneously. For literature review tasks, intrinsic load is determined by the volume and conceptual density of sources, the degree of terminological and methodological heterogeneity across the literature, and the sophistication of the synthesis question. Extraneous load arises from the design of the task environment — cognitive effort expended on processes that do not directly contribute to learning or performance, such as navigating a disorganised document interface, deciphering inconsistent citation formats, or verifying the accuracy of a retrieved claim. Germane load refers to the cognitive effort invested in meaningful schema construction — the building and refinement of mental models that constitute genuine comprehension and can be transferred to novel problems.

The instructional design implication of this taxonomy is that tools and environments should aim to minimise extraneous load and optimise the allocation of remaining cognitive resources toward germane processing — sometimes described as "making room for thinking" (Kirsh, 2013). This principle has been the theoretical engine behind decades of research on multimedia learning, worked examples, split-attention effects, and collaborative learning (Mayer, 2009). Its application to AI-assisted research remains, however, nascent.

4.2 Applying CLT to AI-Assisted Literature Review

Literature review and synthesis tasks impose high intrinsic load by design: a comprehensive review requires simultaneous management of a large number of interacting conceptual elements across diverse sources. The question for AI tool designers is therefore not whether intrinsic load can be eliminated — it cannot, without eliminating the task itself — but whether the extraneous demands of the research environment can be reduced without inadvertently displacing germane elaboration.

Open-ended LLMs like ChatGPT introduce a novel category of extraneous load not present in unaided literature review: verification uncertainty. When a researcher cannot rely on the model's factual accuracy, every synthesised claim requires independent verification — a demand that may substantially exceed the cognitive cost of original source reading. Consistent with this analysis, Doshi and Hauser (2023) found in a controlled experiment that researchers using an open-ended LLM for ideation tasks reported higher frustration — a reliable correlate of extraneous load (Hart & Staveland, 1988) — than researchers working unaided, despite producing outputs of comparable quality. The authors attributed this counterintuitive finding to verification demands rather than generation difficulty.

Source-grounded tools like NotebookLM theoretically reduce verification uncertainty by constraining outputs to cited user sources. If this constraint is effective, it should reduce the extraneous load associated with source checking, freeing cognitive

resources for the germane activity of integrating and evaluating synthesised claims. Whether this theoretical reduction translates into measurable cognitive benefit for academic researchers has not been directly tested. Closest to this question, van Dis et al. (2023) found that academics using citation-linked AI assistance reported lower perceived cognitive effort than those using citation-free assistance, but this study was conducted in a legal research context with a custom tool rather than NotebookLM in an academic setting.

4.3 The Expertise Reversal Effect and Individual Differences

A well-documented moderator of cognitive load effects is learner expertise. The expertise reversal effect (Kalyuga et al., 2003) describes the phenomenon whereby instructional supports that reduce load for novices become redundant or even counterproductive for experts — who have already developed efficient schemas that render scaffolding unnecessary. Applied to AI tool use in research, this predicts that the cognitive benefits of source-grounded assistance should be greatest for researchers who are unfamiliar with a literature domain (acting as relative novices with respect to that corpus) and smallest — or potentially negative — for domain experts whose existing schemas enable efficient unaided synthesis.

This prediction has implications for both research design and institutional practice. Studies that sample researchers from their primary areas of expertise may underestimate the cognitive benefits of AI assistance; studies that place researchers in unfamiliar domains may overestimate them. The expertise reversal effect also suggests that blanket institutional policies on AI tool adoption — either mandating or prohibiting their use across all researchers — are likely to be suboptimal; differentiated guidance calibrated to researcher expertise and task novelty may better serve knowledge production goals.

5. AI Tools and Research Originality: Evidence and Concerns

5.1 Conceptualising Originality in AI-Assisted Research

Research originality is a construct of central evaluative importance yet persistent definitional ambiguity (Guetzkow et al., 2004). At the level of the literature review — the task most directly relevant to NotebookLM use — originality encompasses at least three distinct dimensions. Conceptual novelty refers to the identification of ideas, connections, or categories not explicit in any single source — the capacity to see across the literature what individual studies cannot see within themselves. Integrative depth refers to the ability to weave disparate findings into a coherent theoretical narrative that is more than the sum of its parts. Critical perspective refers to the capacity to identify the limitations, contradictions, and unexplored assumptions of the synthesised literature — what Boote and Beile (2005) describe as the hallmark of scholars rather than mere researchers.

These dimensions map onto CLT's load taxonomy in theoretically tractable ways. Integrative depth requires sustained germane elaboration — the construction of schemas that connect multiple sources — and may therefore benefit from cognitive load reduction if AI tools free resources for this activity. Conceptual novelty, by contrast, requires access to knowledge beyond the immediate corpus, suggesting that closed-corpus tools like NotebookLM may constrain this dimension by limiting the conceptual raw material available for novel synthesis. Critical perspective requires metacognitive distance from the synthesised material — the ability to step back from the corpus and evaluate it from an external standpoint — a capacity that may be supported or undermined by AI assistance depending on whether the tool reinforces or challenges the researcher's existing understanding.

Proposed Operationalization for Future Research: To move beyond conceptual ambiguity, this paper proposes that 'Integrative Depth' be operationally defined as the number and quality of cross-source syntheses that identify convergent or divergent themes not explicitly stated in the source metadata. Furthermore, 'Originality' in the context of the GRCR model should be measured through a combination of semantic divergence (calculating the distance between the AI’s retrieved text and the researcher’s final synthesis) and expert-blind reviews that assess the 'critical independence' of the scholarly narrative.

5.2 Empirical Evidence on AI and Creative Originality

Empirical research on AI's effects on human creativity has produced a complex and somewhat contradictory picture. Studies in creative writing contexts have found both positive and negative effects depending on the mode of AI assistance and the originality dimension assessed. Chan and Zliobaite (2023) found that AI writing assistance reduced the variance of creative outputs across participants — suggesting a homogenising effect that may undermine originality at the population level even when individual outputs improve in technical quality. This "AI homogenisation" concern is particularly salient for scientific originality, where diversity of conceptual approaches is essential for long-term knowledge advancement.

Organisciak et al. (2023) found that AI augmentation can scaffold divergent thinking — the generation of multiple, diverse responses to open-ended problems — particularly for participants who were otherwise constrained by low domain knowledge. This finding is consistent with a "cognitive prosthetic" account in which AI assistance enables researchers to reach originality levels they

would not otherwise achieve, rather than simply replacing or diluting their own originality. The prosthetic account and the homogenisation account are not mutually exclusive: AI assistance may simultaneously lift the floor of originality (benefiting lower-performing researchers) and lower the ceiling (constraining higher-performing researchers), producing a convergence effect.

In strictly academic contexts, Noy and Zhang (2023) conducted a rigorous experiment with knowledge workers and found that access to ChatGPT improved both the speed and rated quality of analytical writing tasks. Critically, quality gains were concentrated in the domain of structure and coherence — dimensions corresponding to integrative depth — rather than novel idea generation, consistent with the theoretical prediction that open-ended LLMs facilitate synthesis more than conceptual novelty. No analogous study has been conducted specifically for source-grounded systems in literature review tasks.

5.3 The Prompting Quality Dimension

A dimension of AI-assisted research originality that has received insufficient empirical attention is the role of researcher prompting — the quality, specificity, and intellectual ambition of the queries posed to the AI system. Preliminary investigations suggest that prompt quality may be as consequential for output originality as any intrinsic property of the AI tool itself. White et al. (2023) conducted a controlled study of prompt variation in LLM-assisted academic summarisation and found that researchers who posed higher-order, evaluative prompts (e.g., "What are the key theoretical tensions in this literature?") obtained outputs rated significantly more original than those who posed lower-order, descriptive prompts (e.g., "Summarise this paper"). This finding has direct implications for the study of NotebookLM, whose interface actively encourages query formulation through its Suggested Questions feature — raising the question of whether the tool's prompting scaffolds constrain or expand the epistemic quality of researcher queries.

The prompting quality dimension also introduces a methodological complexity for research on AI originality effects: the observed originality of AI-assisted outputs conflates the contributions of the tool's architecture, the researcher's domain expertise, and the researcher's prompting skill. Disentangling these contributions requires experimental designs that systematically vary prompting quality while holding tool and researcher characteristics constant — a design challenge that no published study has yet fully addressed.

5.4 Ethical Dimensions: Attribution, Dependence, and Epistemic Autonomy

Beyond empirical questions about originality levels, the use of AI tools in academic research raises normative questions about the nature and attribution of scholarly contribution. If an AI tool identifies a key conceptual tension in a literature that the researcher subsequently develops into a theoretical contribution, to what extent is that contribution the researcher's? Existing academic integrity frameworks — developed for a pre-AI era — offer limited guidance on this question (Lo, 2023; Huang et al., 2023). The specificity of NotebookLM's design adds a further dimension: since the tool's outputs are grounded in the researcher's own uploaded documents, its contributions are constrained to reconfigurations of what the researcher has already gathered — suggesting a closer relationship to research assistance than to independent intellectual contribution.

A second ethical concern centres on cognitive dependence. If researchers regularly rely on NotebookLM for the orientation and synthesis phases of literature review, they may fail to develop the mental models that emerge from direct, unmediated engagement with primary sources. This concern echoes longstanding debates about calculator use in mathematics education and GPS use in spatial navigation — domains where tool reliance has been shown to selectively impair the underlying cognitive skill (Firth, 2020; Carr, 2010). Whether analogous dependence effects operate in AI-assisted research is an empirical question that longitudinal studies — absent from the current literature — will be necessary to address.

6. Methodological and Ethical Challenges for Future Research

6.1 Measurement of Cognitive Load in Ecological Research Contexts

Cognitive load is a latent construct that cannot be directly observed. Existing measurement approaches each carry trade-offs that are particularly acute in naturalistic research contexts. Self-report instruments — most prominently the NASA Task Load Index (Hart & Staveland, 1988) and the Paas Mental Effort Rating Scale (Paas, 1992) — are easy to administer and ecologically flexible, but susceptible to response biases and retrospective distortion. Physiological measures — pupillometry, heart rate variability, EEG-based frontal asymmetry — provide continuous, implicit indices of cognitive load but are intrusive, expensive, and difficult to deploy in authentic research settings. Dual-task paradigms offer a behavioural, objective load index but require secondary task administration that itself alters the primary task context.

For studies of NotebookLM specifically, a further measurement challenge arises from task duration. Authentic literature review tasks extend over days or weeks, whereas most cognitive load studies employ tasks of 30–90 minutes — long enough to

induce genuine load but short enough to complete in a laboratory session. Ecological validity and experimental control are in direct tension, and researchers studying AI-assisted academic work must be explicit about the trade-offs their design choices entail.

Towards a Multi-Method Protocol: Given that authentic research tasks are longitudinal, we suggest a hybrid measurement protocol for studying NotebookLM. This includes: (1) Periodic administration of the NASA Task Load Index (NASA-TLX) to capture subjective mental demand; (2) Process-tracing through screen recording and interaction logs to measure 'verification time' as a proxy for extraneous load; and (3) Post-task 'think-aloud' protocols to identify moments where 'Orientation Scaffolding' successfully transitions into 'Germane Elaboration'.

6.2 Operationalising and Assessing Research Originality

Research originality is even more difficult to operationalise reliably than cognitive load. Existing approaches fall broadly into two categories. Expert rating rubrics assess originality judgements made by domain-knowledgeable evaluators, offering face validity but requiring substantial inter-rater reliability work and being susceptible to disciplinary conventions that may not generalise. Computational text analysis methods — including semantic similarity measures, novelty indices derived from citation networks, and embedding-based divergence scores — offer scalability and replicability but may miss dimensions of originality that depend on domain-specific evaluation standards.

A persistent challenge for originality research in AI-assisted contexts is controlling for baseline originality differences between participants. Researchers who produce highly original outputs with AI assistance may have been highly original without it; without a within-subjects or matched-pairs design, it is impossible to attribute the originality of AI-assisted outputs to the tool rather than to pre-existing researcher capability. Well-controlled studies in this area should employ either within-subjects crossover designs (where the same researcher completes comparable tasks with and without AI assistance) or carefully matched between-subjects designs with baseline originality measures.

6.3 Ethical Considerations in Research on AI-Using Academics

Studies involving academic participants who use AI tools in their research workflows raise several ethical considerations specific to this context. First, participation in a study about AI tool use may itself alter participants' tool use behaviour through demand characteristics — participants may use NotebookLM more or less than they otherwise would because they believe this is what the study expects. Second, studies that require participants to upload their actual research documents to a cloud-based AI system raise data confidentiality concerns, particularly for researchers working with proprietary, sensitive, or unpublished materials. Third, in contexts where institutional AI policies are actively evolving, research participation may inadvertently expose participants to professional risk if their AI use is disclosed to institutional authorities.

7. The Grounded Retrieval–Cognitive Redistribution (GRCR) Model

7.1 Model Overview

Drawing on the theoretical and empirical analysis in Sections 3–6, we propose the Grounded Retrieval–Cognitive Redistribution (GRCR) model as a conceptual framework for predicting and explaining the cognitive and creative effects of source-grounded AI tools in academic research contexts. The model's central claim is that source-grounded tools like NotebookLM do not simply reduce total cognitive load but redistribute cognitive resources across the load taxonomy — reducing extraneous verification load while potentially freeing or potentially constraining germane elaboration, depending on moderating conditions.

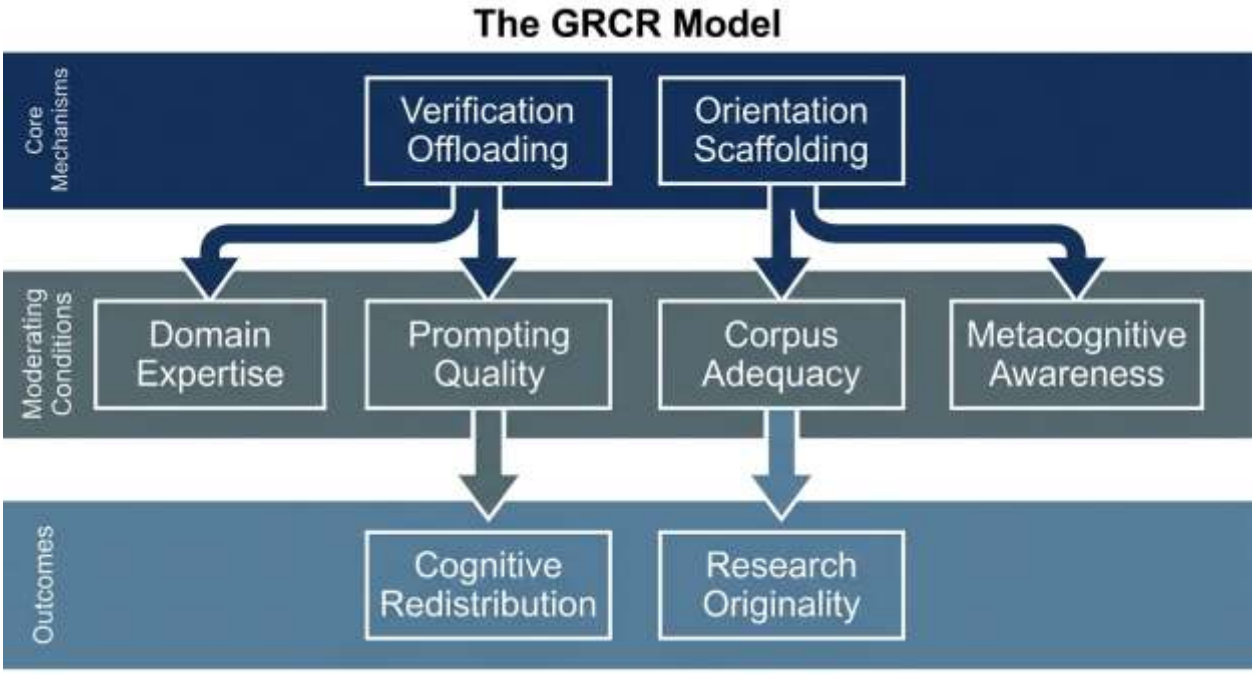

Figure 1. The Grounded Retrieval–Cognitive Redistribution (GRCR) Model, showing the interaction between core mechanisms, moderating conditions, and scholarly outcomes.

7.2 Core Mechanisms

The GRCR model identifies three core mechanisms through which source-grounded retrieval affects cognitive load. The Verification Offloading mechanism describes the reduction in extraneous load attributable to the tool's citation constraints: because outputs are grounded in user-supplied sources and linked to specific passages, researchers can verify claims by consulting cited passages rather than conducting independent searches. The Orientation Scaffolding mechanism describes the reduction in the initial orientation cost of an unfamiliar corpus: NotebookLM's automatic overview generation provides a structured entry point that reduces the effort required to identify the scope and structure of a literature before detailed engagement. The Elaboration Space mechanism describes the cognitive resources made available for germane processing as a consequence of reduced extraneous load: freed from verification and orientation demands, researchers can direct attention toward synthesis, evaluation, and theorisation.

7.3 Moderating Conditions

The GRCR model specifies four conditions that moderate whether freed elaboration space translates into enhanced originality or is displaced by new tool-specific demands. Corpus Adequacy — whether the researcher's uploaded documents comprehensively cover the relevant literature — determines whether the tool's grounding constraint supports or limits the conceptual reach of synthesis. Domain Expertise — the researcher's existing knowledge of the synthesised literature — determines the magnitude of the expertise reversal effect and thus the net cognitive benefit of tool use. Prompting Quality — the intellectual ambition and specificity of the researcher's queries — determines whether the tool's responses stimulate or merely confirm the researcher's existing understanding. Finally, Metacognitive Awareness — the researcher's capacity to recognise and compensate for the tool's limitations — determines whether the tool's constraints (particularly the closed-corpus constraint on conceptual novelty) are experienced as genuine epistemic limitations or as invisible shapers of the synthesis output.

7.4 Model Predictions and Testable Hypotheses

The GRCR model generates several testable predictions. Researchers using NotebookLM should report lower extraneous cognitive load than those using open-ended LLMs or working unaided, particularly on dimensions related to source verification and navigation. The cognitive load advantage of NotebookLM should be moderated by domain expertise, being greatest for researchers in unfamiliar literature domains. Originality gains attributable to NotebookLM use should be concentrated in integrative depth and critical perspective rather than conceptual novelty, consistent with the theoretical constraint that closed-corpus grounding limits cross-domain conceptual reach. Prompting quality should moderate originality outcomes, with high-order prompting producing larger originality benefits regardless of tool condition.

8. Gaps in the Literature and Research Agenda

Our review identifies five critical gaps in the existing literature that future research must address to provide an evidence base adequate for guiding the responsible integration of source-grounded AI tools in academic research.

8.1 Direct Empirical Study of NotebookLM

Despite NotebookLM's growing adoption and distinctive architectural properties, no published peer-reviewed study has directly examined its cognitive or creative effects on academic researchers. Studies of RAG systems in general (Gao et al., 2023; Lewis et al., 2020) and of open-ended LLMs in research contexts (Noy & Zhang, 2023; Liang et al., 2024) provide relevant context but cannot substitute for tool-specific investigation. We call for controlled experiments that directly compare NotebookLM, open-ended LLM assistance, and unaided synthesis on both cognitive load and originality outcomes in academic populations.

The Need for Pilot Case Studies: Given the total absence of peer-reviewed empirical data on NotebookLM as of late 2024, there is an immediate need for documented case studies. Future researchers should implement 'within-subject' designs where academics use NotebookLM for a real-world literature synthesis over several days. Such pilot studies would provide essential qualitative data on whether the tool's 'closed-corpus' nature creates a 'filter bubble' effect, potentially limiting the conceptual novelty that arises from spontaneous, non-grounded analogical reasoning.

8.2 Longitudinal Investigation of Expertise and Dependence

The expertise reversal effect and the cognitive dependence concern — both theoretically significant — require longitudinal designs to investigate properly. Cross-sectional comparisons of high- and low-AI-experience researchers conflate current tool use with pre-existing expertise differences. Longitudinal studies tracking researchers' cognitive load, originality, and tool use patterns over months or years are necessary to determine whether AI tool use builds, maintains, or erodes the underlying cognitive capacities relevant to scholarly synthesis.

8.3 Discipline-Specific Investigation

The cognitive and epistemic effects of source-grounded AI are likely to vary significantly across disciplines with different literature structures, synthesis norms, and originality criteria. Systematic, discipline-comparative research — spanning the natural sciences, social sciences, humanities, and professional fields — is needed to determine the boundary conditions of GRCR model predictions and to develop discipline-specific guidance on appropriate AI tool use.

8.4 Prompting Pedagogy Research

Given the evidence that prompting quality moderates AI output originality, research on how researchers can be trained to formulate higher-order, epistemically ambitious queries — what might be termed prompting pedagogy for researchers — represents a high-priority applied research direction. Such research should examine the effects of structured prompting training on both the quality of AI-assisted outputs and the cognitive demands of the AI interaction.

8.5 Ethical and Policy Frameworks

The ethical dimensions of AI-assisted research — including attribution, dependence, epistemic autonomy, and data confidentiality — require empirical investigation alongside theoretical analysis. Studies documenting how academic communities understand and navigate these dimensions, and how institutional policies shape (or fail to shape) researcher behaviour, are essential for developing governance frameworks that are both principled and practically workable.

9. Conclusion

This narrative review has examined the literature at the intersection of source-grounded AI, Cognitive Load Theory, and academic research originality — three domains that have not previously been brought together in systematic analysis. Our central finding is that the theoretical case for source-grounded tools like NotebookLM conferring cognitive and creative advantages in academic research is coherent and well-grounded in adjacent evidence, but has not yet been empirically tested in directly relevant populations and contexts. The GRCR model proposed in Section 7 offers a theoretical framework for guiding that empirical programme, identifying the mechanisms through which grounded retrieval may redistribute cognitive resources and the moderating conditions that determine whether this redistribution translates into enhanced scholarly originality.

The implications of this review extend beyond any single tool. As source-grounded AI systems become increasingly sophisticated and widely adopted, understanding the conditions under which they support or substitute for researchers' own cognitive and creative contributions becomes a matter of considerable importance for research quality, research training, and the long-term health of academic knowledge production. The literature we have reviewed suggests that the relationship between AI assistance and human cognition is neither uniformly enabling nor uniformly diminishing — it is mediated by task characteristics, individual expertise, tool design, and researcher metacognitive awareness in ways that demand careful, contextually sensitive empirical investigation.

We conclude with a call for interdisciplinary collaboration — between cognitive psychologists, information scientists, AI researchers, and domain scholars — to build the evidence base that institutions, researchers, and tool designers need to navigate the AI-assisted research landscape responsibly and productively.

References

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922
2. Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. Educational Researcher, 34(6), 3–15. https://doi.org/10.3102/0013189X034006003
3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
4. Carr, N. (2010). The shallows: What the internet is doing to our brains. W. W. Norton & Company.
5. [5] Chan, C. K. Y., & Zliobaite, I. (2023). AI assistance reduces the originality of creative writing. arXiv preprint arXiv:2309.05010.
6. Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24(1), 87–114. https://doi.org/10.1017/S0140525X01003922
7. Doshi, A. R., & Hauser, O. (2023). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28). https://doi.org/10.1126/sciadv.adn5290
8. Ferrari, R. (2015). Writing narrative style literature reviews. Medical Writing, 24(4), 230–235. https://doi.org/10.1179/2047480615Z.000000000329
9. Barhoom, Alaa, et al., “Sarcasm Detection in Headline News using Machine and Deep Learning Algorithms”, International Journal of Engineering and Information Systems (IJEAIS), vol. 6, no. 4, pp. 66-73, 2022.
10. Elsharif, Abeer Abed ElKareem Fawzi, et al., “Retina Diseases Diagnosis Using Deep Learning”, International Journal of Academic Engineering Research (IJAER), vol. 6, no. 2, pp. 11-37, 2022.
11. Abunasser, Basem S, et al., “Breast Cancer Detection and Classification using Deep Learning Xception Algorithm”, International Journal of Advanced Computer Science and Applications, vol. 13, no. 7, 2022.
12. Alghoul, Asil Mustafa, et al., “Predictive Analysis of Lottery Outcomes Using Deep Learning and Time Series Analysis”, International Journal of Engineering and Information Systems (IJEAIS), vol. 7, no. 10, pp. 1-6, 2023.
13. Abu-Jamie, Tanseem N, et al., “Classification of Sign-Language Using MobileNet-Deep Learning”, International Journal of Academic Information Systems Research (IJAISR), vol. 6, no. 7, pp. 29-40, 2022.
14. Obaid, Tareq, et al., “Age and Gender Classification from Retinal Fundus Using Deep Learning”, pp. 171-180, 2022.
15. Abunasser, Basem S, et al., “Predicting Customer Revenue in E-commerce Using Machine Learning a Case Study of the Google Merchandise Store”, pp. 27-38, 2023.
16. Salman, Fatima Maher, et al., “Classification of Real and Fake Human Faces Using Deep Learning”, International Journal of Academic Engineering Research (IJAER), vol. 6, no. 3, pp. 1-14, 2022.
17. Abu-Naser, Samy S, et al., “Heart Disease Prediction Using a Group of Machine and Deep Learning Algorithms”, pp. 181-196, 2022.
18. Almzainy, Mohammed, et al., “Classification of Pineapple and Mini Pineapple Using Deep Learning: A Comparative Evaluation”, International Journal of Academic Information Systems Research (IJAISR), vol. 9, no. 1, pp. 23-27, 2025.
19. Kassab, Mohammed Khair I, et al., “Image-Based Tea Leaves Diseases Detection Using Deep Learning”, International Journal of Academic Information Systems Research (IJAISR), vol. 9, no. 6, 2025.
20. Abunasser, Basem S, et al., “Prediction of Instructor Performance using Machine and Deep Learning Techniques”, International Journal of Advanced Computer Science and Applications, vol. 13, no. 7, 2022.
21. Dheir, Ibtesam M, et al., “Classification of Anomalies in Gastrointestinal Tract Using Deep Learning”, International Journal of Academic Engineering Research (IJAER), vol. 6, no. 3, pp. 15-28, 2022.
22. Barhoom, Ali MA, et al., “Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms”, International Journal of Engineering and Information Systems (IJEAIS), vol. 6, no. 4, pp. 1-13, 2022.
23. Ihlayyel, Mazen SM, et al., “Detection and Classification of Tomato Leaf Diseases Using Deep Learning”, International Journal of Academic Information Systems Research (IJAISR), vol. 9, no. 6, pp. 21-28, 2025.
24. Abu Naser, Samy S, et al., “Trends of Palestinian Higher Educational Institutions in Gaza Strip as Learning Organizations”, International Journal of Digital Publication Technology, vol. 1, no. 1, pp. 1-42, 2017.
25. Almzainy, Mohammed S, et al., “Detection and Classification of Faked and Genuine Money Using Deep Learning”, pp. 237-248, 2024.
26. Alkahlout, Mohammed A, et al., “Advances in Kidney Cancer Detection: Harnessing the Power of Deep Learning”, vol. 1, pp. 251, 2025.
27. Alkahlout, Mohammed A, et al., “Advances in Kidney Cancer Detection: Harnessing the Power of Deep Learning for Accurate Diagnosis”, pp. 251-263, 2024.
28. AlDaya, Duha KA, et al., “Predicting Smoking-Associated Thyroid Dysfunction Using Explainable Machine Learning”, International Journal of Academic Engineering Research (IJAER), vol. 9, no. 12, pp. 97-105, 2025.
29. Abu-Naser, Samy S, et al., “Predicting Whether Student will continue to Attend College or not using Deep Learning”, International Journal of Engineering and Information Systems (IJEAIS), vol. 6, no. 6, pp. 33-45, 2022.
30. Alsaqqa, Azmi H, et al., “Detecting Cybersecurity Threats Using Convolutional Neural Networks and Machine Learning”, pp. 15-27, 2025.
31. Dwimah, Amal, et al., “Symbolic Hybrid Knowledge-Based Systems: Integrating Knowledge-Based Reasoning and Machine Learning in Explainable AI”, International Journal of Academic Engineering Research (IJAER), vol. 9, no. 8, pp. 80-84, 2025.
32. Abunasser, Basem S, et al., “Predicting Stock Prices using Artificial Intelligence: A Comparative Study of Machine Learning Algorithms.”, International Journal of Advances in Soft Computing & Its Applications, vol. 15, no. 3, 2023.
33. Taha, Aya Helmi Abu, et al., “Predicting Loan Defaulters: A Comprehensive Analysis and Comparative Study of Machine Learning Algorithms Using a Large-Scale Loan Default Dataset”, pp. 15-28, 2025.
34. Taha, Ashraf MH, et al., “Exploring Emotion Recognition Through EEG Brainwave Data: A Comparative Analysis of Machine Learning and Deep Learning Approaches”, pp. 77-89, 2025.
35. Abdalmenem, Samia AM, et al., “Increasing the Efficiency of Palestinian University Performance through the Implementation of E-Learning Strategies”, International Journal of Academic Management Science Research (IJAMSR), vol. 3, no. 7, pp. 15-28, 2019.
36. Aldaya, Salah Aldin Shadi, et al., “TLEABLCNN: Brain Disorder Detection and Alzheimer’s Using Explainable Attention-Based Deep Learning and SMOTE on Imbalanced MRI Data”.
37. MAlnajjar, Mohammad Khaleel, et al., “Heart Sounds Analysis and Classification for Cardiovascular Diseases Diagnosis using Deep Learning”, International Journal of Academic Engineering Research (IJAER), vol. 6, no. 1, pp. 7-23, 2022.
38. Abu-Naser, Aya Helmi Abu Taha ans Samy S., “Predicting Kian Defaulters: A comprehensive Analysis and Comparative Study of Machine Learning Algorithms Using Dataset, Large-Scale Loan Default”, vol. 3, pp. 15, 2025.
39. Elmahmuom, Abedeleilah Salem Ayyad, et al., “Comparative Analysis of Data Balancing Techniques in Prostate Cancer Classification Using Machine Learning and Deep Learning”, pp. 133-144, 2025.
40. Abu-Naser, Samy S., et al., “Knowledge management in ESMDA: expert system for medical diagnostic assistance”, Artificial Intelligence and Machine Learning Journal, vol. 10, no. 1, pp. 31-40, 2010.
41. Arqawi, Samer M, et al., “Predicting employee attrition and performance using deep learning”, Journal of Theoretical and Applied Information Technology, vol. 100, no. 21, pp. 6526-6536, 2022.
42. Arqawi, Samer M, et al., “Predicting university student retention using artificial intelligence”, International Journal of Advanced Computer Science and Applications, vol. 13, no. 9, 2022.
43. Elzamly, Abdelrafe, et al., “Predicting Software Analysis Process Risks Using Linear Stepwise Discriminant Analysis: Statistical Methods”, Int. J. Adv. Inf. Sci. Technol, vol. 38, no. 38, pp. 108-115, 2015.
44. Mattar, Mohammad S, et al., “Predicting COVID-19 Using JNN”, International Journal of Academic Engineering Research (IJAER), vol. 7, no. 10, pp. 52-61, 2023.
45. ALMASRI, ABDELBASET R, et al., “Predicting instructor performance in higher education using stacking and voting ensemble techniques”, Journal of Theoretical and Applied Information Technology, vol. 103, no. 2, 2025.
46. Husien, Ahmed Mohammed, et al., “Predicting Students' end-of-term Performances using ML Techniques and Environmental Data”, International Journal of Academic Information Systems Research (IJAISR), vol. 7, no. 10, pp. 19-25, 2023.
47. Meqdad, Yousef Mohammed, et al., “Predicting Carbon Dioxide Emissions in the Oil and Gas Industry”, International Journal of Academic Information Systems Research (IJAISR), vol. 7, no. 10, pp. 34-40, 2023.
48. Abu Nada, AM, et al., “Age and Gender Prediction and Validation through Single User Images Using CNN. 2020”.
49. Al Fleet, Muhannad Jamal Farhan, et al., “Predicting Player Power In Fortnite Using Just Nueral Network”, International Journal of Engineering and Information Systems (IJEAIS), vol. 7, no. 9, pp. 29-37, 2023.
50. Samra, Mohammed Nafez Abu, et al., “ANN Model for Predicting Protein Localization Sites in Cells”, International Journal of Academic and Applied Research (IJAAR), vol. 4, no. 9, pp. 43-50, 2020.
51. Bakr, Mohammed Abdul Hay Abu, et al., “Breast Cancer Prediction using JNN”, International Journal of Academic Information Systems Research (IJAISR), vol. 4, no. 10, pp. 1-8, 2020.
52. Al-Madhoun, Osama Salah El-Din, et al., “Low Birth Weight Prediction Using JNN”, International Journal of Academic Health and Medical Research (IJAHMR), vol. 4, no. 11, pp. 8-14, 2020.
53. Oriban, Ahmed Jabara Abu, et al., “Antibiotic Susceptibility Prediction Using JNN”, International Journal of Academic Information Systems Research (IJAISR), vol. 4, no. 11, pp. 1-6, 2020.
54. Shawarib, Mohammed Ziyad Abu, et al., “Breast Cancer Diagnosis and Survival Prediction Using JNN”, International Journal of Engineering and Information Systems (IJEAIS), vol. 4, no. 10, pp. 23-30, 2020.
55. Alkayyali, Zakaria KD, et al., “Comparative Analysis of Regressor Models for Predicting Heart Attack Risk: A Comprehensive Evaluation Using Regression Metrics and Visualization”, pp. 119-132, 2025.
56. Abu-Jamie, Tanseem N, et al., “Classification of Sign-language Using VGG16”, International Journal of Academic Engineering Research (IJAER), vol. 6, no. 6, pp. 36-46, 2022.
57. Alkayyali, Zakaria KD, et al., “Classification of Cardiovascular ECGs Using MODWPT-Based Feature Extraction: A Comparative Study on Four Ailments from MIT-BIH Databases”, pp. 225-236, 2024.
58. Saad, Abdulrahman Muin, et al., “Rice Classification using ANN”, International Journal of Academic Engineering Research (IJAER), vol. 7, no. 10, pp. 32-42, 2023.
59. Alzamily, Jawad Yousef Ibrahim, et al., “Classification of encrypted images using deep learning–Resnet50”, Journal of Theoretical and Applied Information Technology, vol. 100, no. 21, pp. 6610-6620, 2022.
60. Almassri, Mohammed M, et al., “Grape Leaf Species Classification Using CNN”, International Journal of Academic Information Systems Research (IJAISR), vol. 8, no. 4, pp. 66-72, 2024.
61. Zarandah, Qasem MM, et al., “Efficient Respiratory Disease Classification Using Customized CNN on a Large Kaggle Dataset”, pp. 105-117, 2025.
62. Marouf, Mohammed, et al., “Fine-tuning MobileNetV2 for Sea Animal Classification”, International Journal of Academic Information Systems Research (IJAISR), vol. 8, no. 4, pp. 44-50, 2024.
63. Barhoom, Alaa MA, et al., “A survey of bone abnormalities detection using machine learning algorithms”, vol. 2808, no. 1, pp. 040009, 2023.
64. Abunasser, Basem S, et al., “Literature review of breast cancer detection using machine learning algorithms”, vol. 2808, no. 1, pp. 040006, 2023.
65. Abdaljawad, Rabah YR, et al., “Fraudulent financial transactions detection using machine learning”, pp. 1-9, 2023.
66. Taha, Ashraf MH, et al., “Impact of data augmentation on brain tumor detection”, Journal of Theoretical and Applied Information Technology, vol. 101, no. 11, 2023.
67. Abunasser, Basem S, et al., “Convolution neural network for breast cancer detection and classification-final results”, Journal of Theoretical and Applied Information Technology, vol. 101, no. 1, pp. 315-329, 2023.
68. Abunasser, Basem S, et al., “Convolution neural network for breast cancer detection and classification using deep learning”, Asian Pacific journal of cancer prevention: APJCP, vol. 24, no. 2, pp. 531, 2023.
69. Taha, A, et al., “Investigating the effects of data augmentation techniques on brain tumor detection accuracy”, Journal of Theoretical and Applied Information Technology, vol. 101, no. 11, pp. 9, 2023.
70. Buhisi, Nabil I., et al., “Dynamic programming as a tool of decision supporting”, Journal of Applied Sciences Research; www.aensiweb.com/JASR/, vol. 5, no. 6, pp. 671-676, 2009.
71. Abu-Naser, SAMY S, et al., “The miracle of deep learning in the Holy Quran”, J Theor Appl Inf Technol, vol. 101, pp. 17, 2023.
72. Saleh, Ahmad, et al., “Brain tumor classification using deep learning”, pp. 131-136, 2020.
73. Obaid, Tareq, et al., “Factors affecting students’ adoption of e-learning systems during COVID-19 pandemic: A structural equation modeling approach”, pp. 227-242, 2022.
74. El-Habil, Basel Y, et al., “Global climate prediction using deep learning”, J Theor Appl Inf Technol, vol. 100, no. 24, pp. 4824-4838, 2022.
75. Al-Zamily, Jawad Yousef Ibrahim, et al., “A survey of cryptographic algorithms with deep learning”, vol. 2808, no. 1, pp. 050002, 2023.