

TechNews: A Technical Report on Agentic Workflow Automation for Science News Publishing

Byunghyun Ban

Department of Music, Gyungkook National University
NaNa Lab

Corresponding author: bhban@nanalab.kr

Abstract

This technical report presents TechNews, a registered automated science news portal that transforms newly released scientific papers into publishable science news articles through an agentic production workflow. (<https://technews.ai.kr>)

Rather than treating automated journalism as a single text-generation task, TechNews decomposes human content-production work into persistent stages: source collection, freshness filtering, article drafting, translation, formatting, thumbnail generation, publication upload, backend monitoring, and operational maintenance. The system is organized around two agentic components. The first is a reporter agent, Persica, which performs source collection, article generation, translation, formatting, thumbnail handling, upload preparation, and operational reporting. The second is a backend operations agent that maintains service awareness across the publishing backend, including API behavior, database state, media storage, deployment assumptions, diagnostic routines, and operational failure modes. Together, these agents form an agent-operated publishing infrastructure rather than a one-shot article generator. TechNews operates as a registered internet newspaper in the Republic of Korea under registration number 경북아 00881.

management system, and verify that published content appears correctly.

This report presents TechNews as a case study of transforming this human content-production workflow into an agentic publishing system. TechNews is designed to collect scientific papers whose public release timestamps fall within the previous 24 hours, convert selected sources into science news articles, generate or attach editorial thumbnails, and publish the resulting content through a live production backend.

The novelty of TechNews is not that a model can write article text. Prior work on automated journalism has already examined algorithmic news generation, robotic reporting, and the implications of automation for journalistic labor and authority [1-3]. TechNews extends this discussion from text generation to agentic workflow automation. The system automates not only writing, but the broader path from source discovery to public publication.

This report is therefore best read as a systems white paper and case study rather than as a new-model paper. It does not propose a new language model or benchmark. Instead, it documents how human content-production labor can be decomposed into agent-executable stages and operated in a registered publication environment.

1. Introduction

Large language models have made it increasingly easy to generate fluent article-like text. However, news production is not only a writing task. A functioning news publication must monitor sources, select timely items, interpret source material, prepare headlines and article bodies, handle multilingual formatting, attach visual material, publish through a backend or content-

2. Background and Positioning

2.1 Automated journalism

Automated journalism is usually discussed in relation to algorithmically generated stories, particularly in domains where structured data are abundant. Earlier

research emphasized the potential for automated systems to generate news at scale, reduce production cost, and support multilingual output, while also raising questions about labor, editorial authority, transparency, and accountability [1-3].

TechNews is related to automated journalism, but it is positioned differently. Many automated journalism systems focus on producing a story from a data input. TechNews focuses on the full production workflow: source monitoring, article transformation, visual preparation, publication, and operational support. The unit of automation is not only the generated article, but the path by which a source becomes a published media artifact.

2.2 LLM agents and workflow automation

TechNews also relates to recent work on LLM-based agents. Tool-using and action-oriented LLM frameworks show that language models can be embedded in workflows where they interact with external environments, tools, or state [4]. Surveys of LLM-based agents frame agents as systems that perceive environments, make decisions, and act through tools or interfaces [5].

This framing is important for TechNews. The reporter agent does not merely receive a prompt and return an article. It interacts with internal databases, staging records, scheduling logic, media services, upload routines, logs, and backend state. The backend operations agent similarly maintains contextual knowledge about the production service rather than merely answering generic questions about it.

2.3 Grounding and science-news risk

Automated science news production is especially sensitive to grounding. A system that transforms scientific papers into public-facing articles must avoid making claims unsupported by the source paper. Retrieval-augmented generation combines model knowledge with retrieved evidence and has been shown to improve specificity and factuality in knowledge-intensive generation [6]. Self-RAG extends this idea with adaptive retrieval and self-reflection for improving factuality and citation behavior [7].

The need for grounding is reinforced by the broader hallucination problem in natural language generation [8]. It is also reinforced by evidence from science communication research: health and science news can be distorted by exaggeration, loss of caveats, and

overconfident presentation of uncertain findings [9,10]. For an automated science news portal, source traceability is therefore not a cosmetic feature; it is an editorial requirement.

3. System Context

3.1 Registered science news portal

TechNews is not only a prototype pipeline. It operates as a registered internet newspaper in the Republic of Korea under registration number Gyeongbuk A00881. The public TechNews site displays required publication information such as the title, registration number, registration date, publication date, publisher, editor, and related publication metadata [11].

Korean newspaper law establishes a registration framework for newspapers and internet newspapers, and it also requires internet newspaper operators to publish certain necessary information such as the title, registration number, registration date, publisher, and editor [12]. This regulatory context distinguishes TechNews from a laboratory-only demonstration. However, registration should not be interpreted as governmental certification of the AI system's accuracy, editorial quality, factuality, or reliability.

3.2 Freshness criterion

TechNews uses a freshness criterion based on public release time. A scientific paper is eligible for collection when its public release timestamp falls within the 24-hour window preceding the collector execution time. In this report, a “24-hour paper” refers to this operational collection rule, not necessarily to a formal journal publication date.

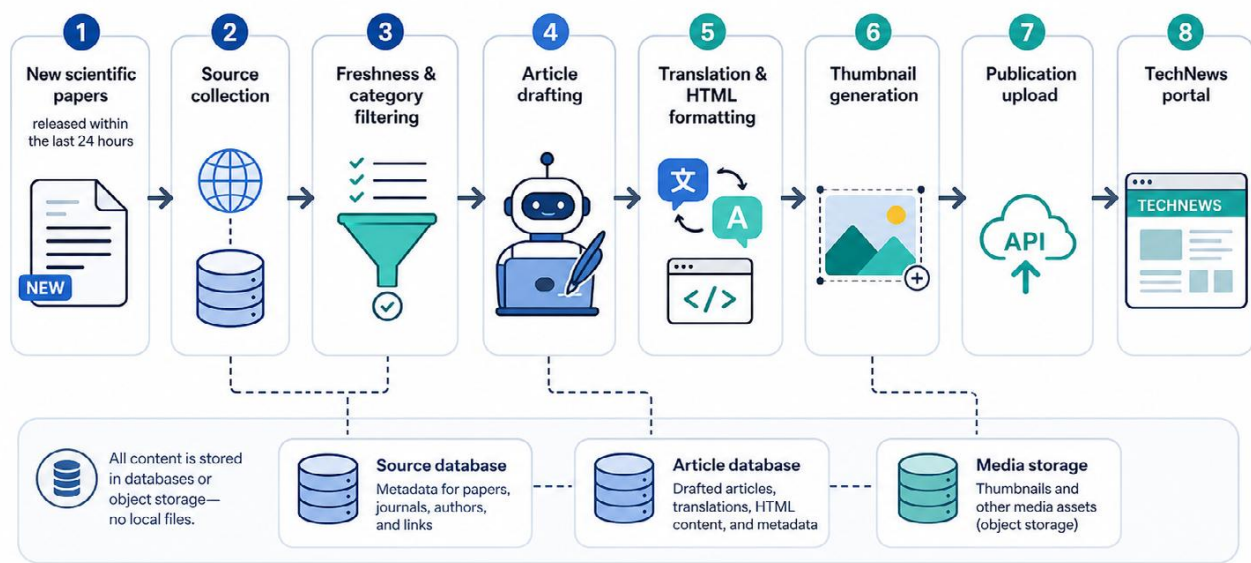


Figure 1. TechNews workflow overview

This distinction matters because many scientific sources expose new papers through public feeds, preprint servers, or API timestamps before conventional journal publication. For example, arXiv provides API and RSS/Atom feed mechanisms for programmatic or feed-based access, while bioRxiv and medRxiv provide API endpoints for retrieving records by date interval or recent-post criteria [13-15].

3.3 Human-governed automation

TechNews is automated at the workflow level, but it should not be described as ungoverned. The system automates collection, drafting, translation, thumbnail handling, publication upload, and backend monitoring, while the human operator remains responsible for policy, correction, intervention, and publication accountability.

This distinction is consistent with broader newsroom guidance on AI. The Associated Press states that AI is not a replacement for journalists and emphasizes human responsibility for accuracy and fairness, while Reuters states that primarily or solely AI-generated news content should be transparently disclosed with context about the role of AI [16,17].

4. Overall Architecture

TechNews consists of three interacting layers: the public news service, the reporter-agent workflow, and the backend operations-agent workflow. Figure 2 summarizes this dual-agent publishing architecture.

The public news service is the publication surface. It serves articles, sections, media, and user-facing pages. The backend environment is treated as a production service stack composed of a web frontend, an API layer, article databases, media storage, and operational routines. Implementation details are intentionally abstracted in this report; the important point is that publication requires a persistent service environment rather than merely a text-generation endpoint.

The reporter agent, Persica, is the editorial production component. It collects source materials, creates article drafts, translates and formats content, generates or reconciles thumbnails, and prepares completed articles for publication. Its specification describes it as a persistent research-and-publishing agent with scheduled work, staged processing, model-assisted generation, thumbnail handling, guarded publication workflows, and operational observability [18].

The backend operations agent is the infrastructure-awareness component. It maintains knowledge about the publishing backend, deployment assumptions, database behavior, service routing, operational scripts,

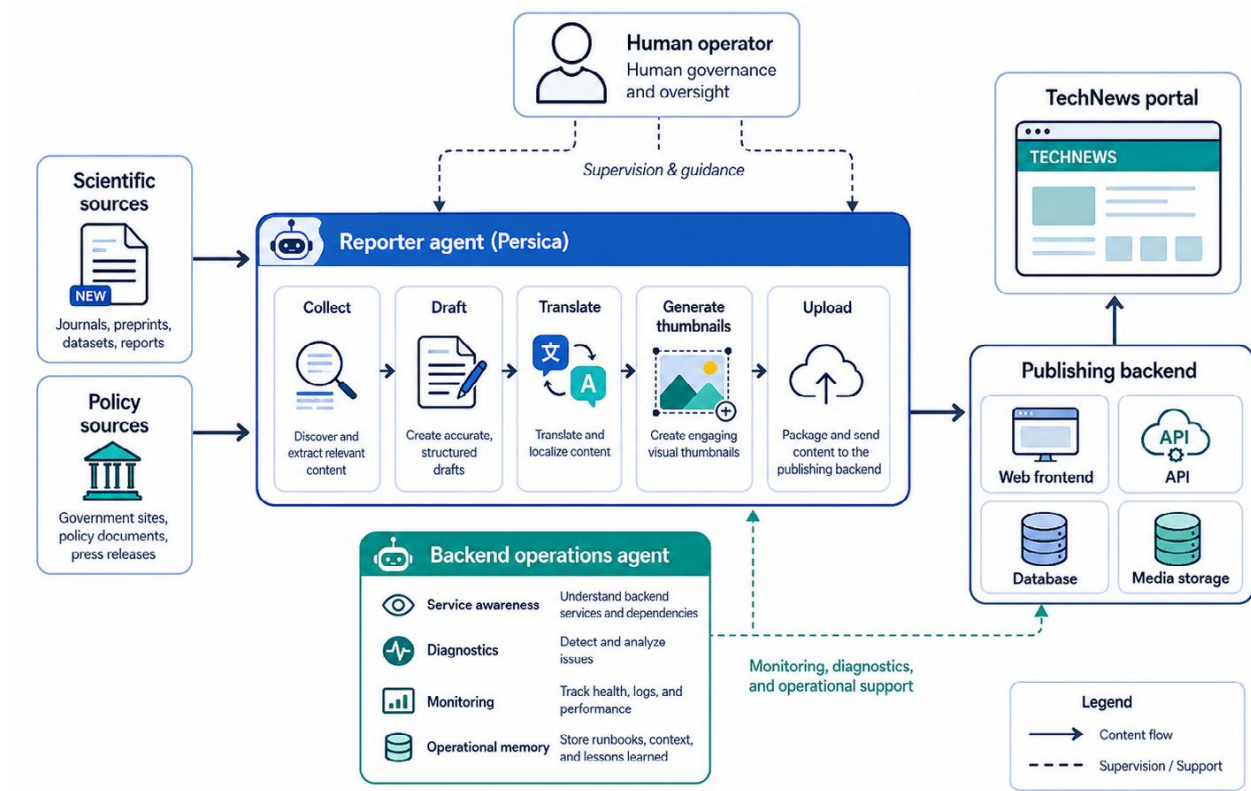


Figure 2. Dual-agent publishing architecture

and failure diagnosis procedures. Its specification describes a digital operations assistant that combines long-term memory, tool-based execution, safety rules, service diagnosis, and backend orchestration [19].

This architecture reflects a central design principle: the article is the output, but the workflow is the system. The generated article matters, but so do the state transitions that lead to it: collection, filtering, drafting, translation, media preparation, upload, logging, and monitoring.

5. Agentic Content-Production Workflow

Science news production is traditionally a human-centered workflow. A human reporter or editor monitors newly released papers, selects potentially relevant items, interprets source material, writes an article, prepares a headline, attaches visual material, publishes the story, and verifies publication state. TechNews decomposes this work into agent-

executable stages. Figure 3 frames this transformation from human editorial labor to an agentic workflow.

Table 2. Human editorial labor mapped to TechNews workflow stages.

Human content-production task	Agentic workflow stage
Monitoring newly released papers	24-hour source collection
Selecting relevant items	Freshness and category filtering
Drafting the article	Model-assisted article drafting
Editing and formatting	Translation and HTML formatting
Preparing visuals	Thumbnail generation or reconciliation
Publishing through CMS or backend	API-based publication upload
Checking publication state	Logs, ledgers, and monitoring records
Handling operational issues	Backend operations-agent diagnosis and support

Table 1. Core system components.

Component	Role	Key responsibility
Reporter agent	Editorial production	Collect sources, draft articles, translate, format, prepare thumbnails, and hand off content for upload.
Backend operations agent	Operational support	Maintain service awareness, diagnose backend issues, and support publication infrastructure.
Article databases	Persistent content state	Store source metadata, article drafts, translations, publication metadata, and status information.
Media storage	Visual assets	Store editorial thumbnails and related media assets.
Human operator	Governance	Set policy, intervene in failures, supervise publication boundaries, and manage accountability.

5.1 Source collection and filtering

The collection stage ingests candidate source items from scientific and policy sources. The system applies the public-release freshness criterion, assigns categories, normalizes links, records metadata, and removes duplicates. The output of this stage is not a finished article. It is a structured source record stored in the internal source database.

This separation is important because collection and writing are different forms of work. Collection asks whether a source is eligible and relevant. Writing asks how that source should be transformed into public-facing explanation. Treating them as separate stages makes the system more inspectable and easier to intervene in.

5.2 Article drafting, translation, and formatting

The drafting stage transforms selected source records into Korean science news article drafts. The translation and formatting stage then prepares a bilingual article representation and web-ready body structure. In the abstracted architecture, these intermediate outputs are persisted in the article database, where drafting, translation, formatting, and publication status can be inspected independently.

This design prevents article generation, translation, and presentation logic from collapsing into one

monolithic process. It also makes failures easier to localize. If a source is collected but not drafted, the failure is different from a translation failure or a publication-upload failure.

5.3 Thumbnail generation and media handling

Science news articles often require visual material, but newly released papers may not provide clean, usable publication thumbnails. Some sources have no image, only small icons, malformed URLs, or figures unsuitable for editorial display. TechNews handles this through a media workflow that generates or attaches editorial thumbnails and stores them in media storage.

Generated images are treated as editorial thumbnails, not as scientific figures or experimental evidence from the original papers. This distinction is essential. A generated thumbnail may make an article more readable, but it must not be presented as a figure from the paper or as evidence of the paper's experimental results. Responsible synthetic media practices also emphasize the need for clear disclosure, provenance, and responsible publishing norms for AI-generated media [20].

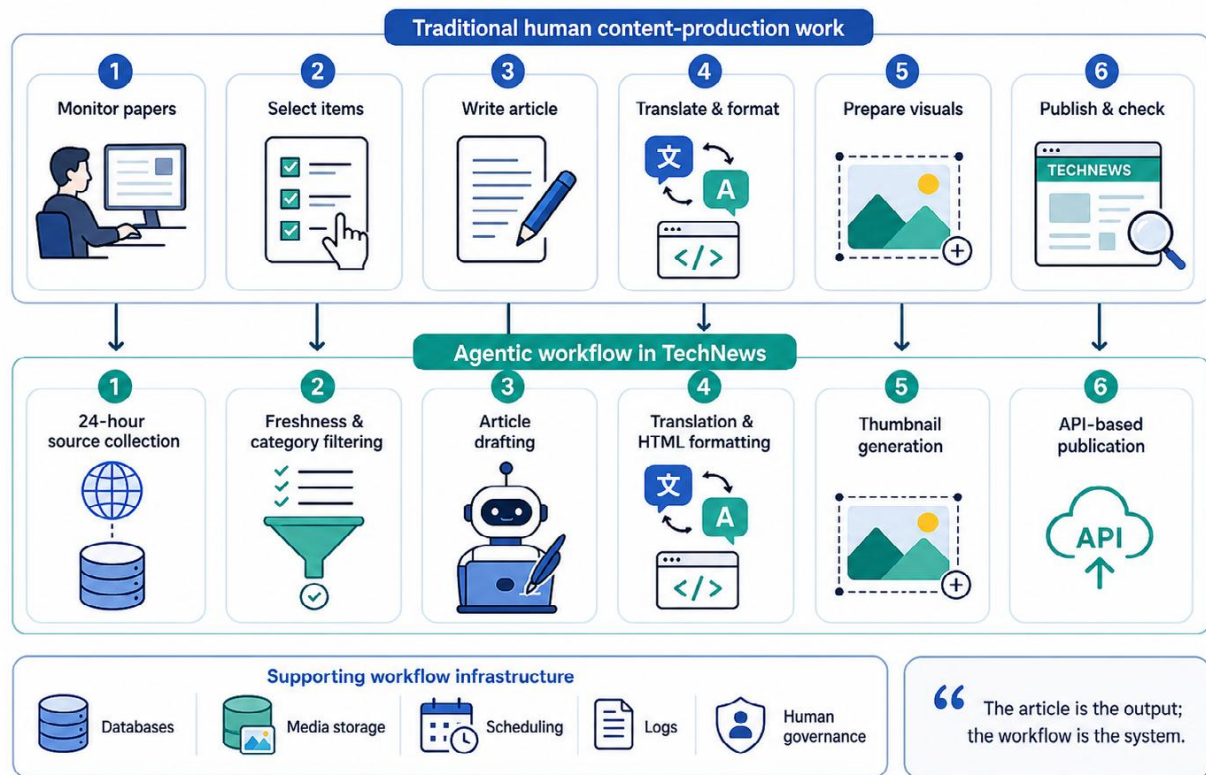


Figure 3. From human editorial labor to agentic workflow

5.4 Publication upload

The upload stage is the publication boundary. Before upload, generated content is an internal artifact. After upload, it becomes a public media product. For that reason, publication is separated from drafting, translation, and thumbnail generation.

At this boundary, the system packages the article, metadata, translations, and media references into a publication payload. Successful publication updates publication records, while failures are retained as operational evidence for later diagnosis. This boundary is the highest-risk point in the workflow because an internal generation becomes public content.

6. Operations, Observability, and Reliability

6.1 Database-backed state and observable transitions

TechNews treats workflow state as an externalized operational artifact. In this report, those artifacts are described as internal databases, staging records, and storage layers. Source records, draft articles, translated articles, media assets, upload status, failure records, and operational summaries should all be inspectable without relying on the model's conversational memory.

This design makes automation auditable. Instead of asking whether an agent claims to have completed a task, the operator can inspect durable state transitions: source collected, article drafted, translation completed, thumbnail attached, payload uploaded, publication confirmed, or failure recorded.

6.2 Scheduling and worker execution

TechNews is a scheduled workflow system. Source collection, article drafting, translation, thumbnail generation, upload, and reporting can be triggered by scheduled workers. This is a major difference from one-shot chatbot use. The agentic system must continue to operate across time, not merely respond in a single conversation.

Scheduling creates operational challenges. Long-running jobs can be ambiguous if logs are delayed or if a worker is terminated. Shorter, staged workers improve observability because each execution has a smaller scope and clearer expected outcome. The Persica specification records several operational changes motivated by this issue, including replacing long draining runs with shorter repeated runs and adding stronger status reporting [18].

6.3 Reliability risks

Automated publishing reliability is not only a model-quality problem. It is also a workflow, state, scheduling, and operations problem. Representative risks include duplicate ingestion, source misclassification, stale worker state, overlapping scheduled tasks, partial media generation, upload failures, and backend drift.

A particularly important class of failures occurs at the publication boundary. If a system treats an upload as successful before durable confirmation, an article may be lost or duplicated. If an upload limiter is misconfigured, a small manual test can accidentally become a batch publication. The Persica specification records this type of operational bug and its remediation as a concrete lesson from live use [18].

Backend drift is another risk. A publishing workflow depends on routes, database schema, media paths, frontend rendering, and deployment state. The backend operations-agent specification emphasizes that a production publishing system may have multiple article tables, views, service routes, and deployment paths that must be understood as an operational whole [19].

6.4 Design lesson

The design lesson is that agentic media automation must make state transitions visible. Logs, database records, media objects, run receipts, and failure records are not peripheral. They are the evidence that the workflow actually occurred. A fluent report from

an agent is weaker evidence than a traceable state transition.

7. Governance and Limitations

7.1 Editorial responsibility

TechNews demonstrates that substantial content-production work can be automated, but automation does not remove responsibility. A registered news portal must remain accountable for the content it publishes. Automated source selection, article drafting, translation, and visual generation all require policy constraints and human-governed oversight.

For science news, this responsibility is especially important. Scientific papers, particularly preprints, may be preliminary, domain-specific, uncertain, or not yet peer reviewed. Automated articles should avoid overstating findings, presenting generated thumbnails as scientific evidence, or implying certainty not supported by the source.

7.2 Disclosure and representation boundaries

The report treats three boundaries as central: source-grounding, synthetic-media representation, and publication. Source-grounding concerns whether article claims remain attributable to the source material. Synthetic-media representation concerns whether generated thumbnails are clearly understood as editorial illustrations rather than scientific figures. Publication concerns the moment at which internal content becomes public media.

These boundaries are not solved by automation alone. They require editorial policy, disclosure norms, correction processes, and human accountability. This is consistent with guidance from news organizations that AI may assist newsroom processes but does not remove responsibility for accuracy, fairness, and transparency [16,17].

7.3 Limitations

- This is a single-system technical report. It does not claim that TechNews is the only or best architecture for automated science news publishing.
- This draft does not yet include quantitative operational evaluation, such as average daily

source count, publication rate, failure rate, latency from source release to publication, or correction rate.

- The system's reliability depends on source quality, model quality, backend stability, and the quality of operational guardrails.
- Registration as an internet newspaper establishes a legal publication context, but it does not certify the accuracy, factuality, reliability, or quality of the automated system.
- Future versions should include stronger source-grounding checks, claim-level attribution, durable event logs, editorial review dashboards, and post-publication correction workflows.

8. Discussion and Conclusion

TechNews shows that automated science news publishing should be understood as agentic workflow automation rather than article generation alone. The visible product is a published article, but the system's real contribution is the observable workflow that converts newly released scientific sources into public media artifacts.

This distinction changes how automated journalism should be evaluated. A model-generated article can be judged by fluency, accuracy, and readability. A publishing workflow must also be judged by source traceability, state management, upload control, failure recovery, observability, backend reliability, correction mechanisms, and governance.

TechNews also shows why backend awareness matters. A publishing agent that writes articles but does not understand the publication infrastructure can fail at the moment generated text becomes public content. By combining a reporter agent with a backend operations agent, TechNews treats publishing as both an editorial and operational process.

The broader lesson is that future AI media systems may not take the form of a single artificial reporter. They may instead become agentic production environments: systems in which collection, writing, translation, image handling, upload, monitoring, and infrastructure maintenance are coordinated through persistent agents.

References

- [1] A. Graefe, Guide to Automated Journalism. Tow Center for Digital Journalism, Columbia Journalism School, 2016.
- [2] M. Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority," *Digital Journalism*, vol. 3, no. 3, pp. 416-431, 2015.
- [3] N. Diakopoulos, *Automating the News: How Algorithms Are Rewriting the Media*. Harvard University Press, 2019.
- [4] S. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," *arXiv:2210.03629*, 2022.
- [5] Z. Xi et al., "The Rise and Potential of Large Language Model Based Agents: A Survey," *arXiv:2309.07864*, 2023.
- [6] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *NeurIPS*, 2020.
- [7] A. Asai et al., "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection," *arXiv:2310.11511*, 2023.
- [8] Z. Ji et al., "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys*, 2023.
- [9] P. Sumner et al., "The association between exaggeration in health related science news and academic press releases," *BMJ*, 2014.
- [10] F. van Schalkwyk et al., "Reporting preprints in the media during the COVID-19 pandemic," *Public Understanding of Science*, 2022.
- [11] TechNews public site, required publication information, accessed May 2026.
- [12] Republic of Korea, Act on the Promotion of Newspapers, etc., Articles 9 and 21.
- [13] arXiv, arXiv API User's Manual.
- [14] arXiv, RSS and Atom feed documentation.
- [15] bioRxiv/medRxiv API documentation.
- [16] Associated Press, "Standards around generative AI," 2023.
- [17] Reuters, "Reuters and AI," accessed May 2026.
- [18] Author's internal technical specification, Persica reporter agent system, 2026.
- [19] Author's internal technical specification, backend operations agent system, 2026.
- [20] Partnership on AI, Responsible Practices for Synthetic Media: A Framework for Collective Action, 2023.